Predicting Equity Risk Premium: Conditioning Forecasts on Economic Uncertainty*

V. Ravi Anshuman†

Rajdeep Sharma‡

June 29, 2025

*We acknowledge the helpful comments on this paper from Ravi Jagannathan, Bidisha Chakrabarty, Partha Mohanram, Srinivasan Rangan, and seminar participants at the International Conference on Financial Markets and Corporate Finance 2022, the NYCU International Finance Conference 2022, the New Zealand Finance Meeting 2022, the Paris Financial Management Conference 2022, the Emerging Markets Conference 2022, the India Finance Conference 2022, the Financial Markets and Corporate Governance Conference 2023, the International Finance and Banking Society Conference 2023, and the International Risk Management Conference 2023. We thank Prateek Jain and Arnab Biswas for providing excellent research assistance. All remaining errors are our own. Send correspondence to V. Ravi Anshuman, Indian Institute of Management, Bangalore, Bannerghatta Main Road, Bengaluru, Karnataka 560076, India. E-mail: anshuman@iimb.ac.in.

†Indian Institute of Management, Bangalore

‡S.P. Jain Institute of Management and Research, Mumbai

Predicting Equity Risk Premium: Conditioning Forecasts on Economic Uncertainty

Abstract

Prior research has shown that individual predictors of the market equity risk premium fail to deliver superior out-of-sample forecast accuracy relative to the historical average. Combining forecasts of individual predictors also underperforms the historical average in recent decades. In this article, we construct an optimal combination of predictors across two different regimes of economic uncertainty (normal and high) and show that this optimal combination delivers the highest out-of-sample forecast accuracy. Specifically, we find that the earnings-price ratio has the lowest forecast error during periods of normal economic uncertainty, while inflation has the lowest forecast error during periods of high economic uncertainty. Investors can use our economic uncertainty-conditioned predictor combination to generate more accurate forecasts of the equity risk premium; actionable trading strategies based on this approach yield utility gains of around 4% and a haircut Sharpe Ratio of 0.41, which are the largest among all other predictors.

JEL classification: C22, G12, G14

Keywords: Equity risk premium, Out-of-sample, Economic uncertainty, Volatility Index, Utility gain, Sharpe ratio

Introduction

Contemporary empirical asset pricing research has largely been shaped by two agendas: explaining cross-sectional variation in expected returns and understanding the time-series dynamics of the aggregate market equity risk premium. This study contributes to the latter by focusing on the predictability of the market equity risk premium (ERP), a key variable with far-reaching implications for macroeconomic outcomes. The estimate of ERP is, therefore, a critical factor in investment decisions of both individual as well as institutional investors. Not surprisingly, considerable research effort has focused on *predicting* ERP. Extant literature has relied on a standard set of economic variables as potential predictors. These predictors reflect information contained in earnings, dividends, cash flow generating capacity, book value, interest rates, macroeconomic indicators, and volatility measures.[1]

However, the predictability of ERP is elusive, as highlighted by Spiegel (2008).[2] Welch and Goyal (2008) provide a comprehensive analysis of the horse-race between different predictors of ERP and find that individual predictors fail to generate superior out-of-sample forecasts, relative to the historical average ERP. They attribute the poor out-of-sample performance of predictors to structural instability, which causes a shift in the return-generating process that underlies equity risk premium.

Motivated by this line of reasoning, Rapach et al. (2010) argue that individual predictors capture different aspects of the complex return-generating process underlying ERP and, therefore, suggest combining forecasts from in-

[1]The relevant literature is as follows: earnings (Campbell and Shiller, 1988a; Lamont, 1998), dividends (Rozeff, 1984; Campbell and Shiller, 1988b; Fama and French, 1988; Goyal and Welch, 2003), cash flow generating capacity (Rayburn, 1986; Hecht and Vuolteenaho, 2006; Westerlund and Narayan, 2014), book value (Kothari and Shanken, 1997; Pontiff and Schall, 1998; Campbell and Shiller, 2001), interest rates (Ball, 1978; Campbell, 1987), macroeconomic indicators (Lintner, 1975; Nelson, 1976; Fama and Schwert, 1977; Fama, 1981), and volatility measures (French et al., 1987; Baillie and DeGennaro, 1990; Campbell and Hentschel, 1992).

[2]See Boudoukh et al. (2008); Campbell and Thompson (2008); Cochrane (2008); Welch and Goyal (2008).

dividual predictors to improve forecast accuracy. Rapach et al. (2010) show that the mean combination forecast – the average of forecasts from individual predictive regression models – outperforms individual predictors and the historical average ERP, in terms of out-of-sample predictability. They argue that combining forecasts delivers superior forecasts because it incorporates information from several economic variables while reducing forecast volatility. Combination forecasts, however, may not deliver superior forecasts if the individual predictors capture similar information on the return-generating process of ERP. In a recently published paper, Denk and Löffler (2024) do not find evidence of the superior predictive ability of combination forecasts in recent decades. They attribute this lack of predictive ability to an increase in the correlation of forecast errors of individual predictive regression models, thereby exacerbating forecast errors of combination forecasts.

Several factors – including economic variables with potential predictability, as well as regime shifts due to changes in investor preferences, institutional reforms, policy shocks, and developments in information technology – can cause an uncertain, complex, and continuously evolving return-generating process for expected equity returns (Timmermann, 2008). Since individual predictors may have poor out-of-sample performance due to dynamic shifts in the economic environment, it is unlikely that individual predictors would be able to consistently generate reliable ERP forecasts over the entire out-of-sample period. What may be more likely is that even the best individual predictors perform better only during particular economic regimes where their influence is strong.

In this paper, we build a prediction framework to capture the dynamic shifts in the economic environment that could influence the first moment (expected return) of the return-generating process of ERP. Specifically, we argue that, by conditioning the forecasts of the predictors on *economic uncertainty*, it is possible to obtain superior forecasts of ERP. This is consistent with Fama and French (1989), Cochrane (1999), and Cochrane (2007) who argue that investors require a higher risk premium during periods of increased risk aversion (e.g., during business cycle downturns). Conversely, investors require a lower risk premium during periods of decreased risk aversion (e.g., during business cycle upturns).

Our procedure relies on using the time-series variation in economic uncertainty to identify regime shifts in the economic environment. We propose an optimal switching algorithm, which recognizes that while the forecast accuracy of a given predictor may be mediocre over the entire out-of-sample period, it may perform much better during subperiods or regimes when economic uncertainty is at a particular level. Thus, our procedure implicitly assumes that the return-generating process of ERP changes across the different economic uncertainty regimes - which implies that it may not be optimal to track the dynamic return-generating process of ERP by employing a *single* predictor over the entire out-of-sample period. Instead, one might do better by adopting a flexible approach of optimally switching between predictors that perform better in different regimes of economic uncertainty.

We measure economic uncertainty by the Volatility Index (VIX), which is a forward-looking measure of economic uncertainty (available from 1990 onward). Although VIX is a widely used measure of economic uncertainty, it has limited linear predictive power for the equity risk premium (Bollerslev et al., 2009). Instead, it serves more effectively as a regime indicator, reflecting shifts in sentiment and macroeconomic uncertainty (Bekaert and Hoerova, 2014). Recent work highlights its nonlinear role: ERP increases mainly when VIX exceeds high thresholds (Bansal and Stivers, 2025), and negative VIX shocks tend to precede higher returns (Dennis et al., 2006). These asymmetric patterns are consistent with broader evidence that uncertainty shocks – proxied in part by VIX – play a central role in driving asset price dynamics, particularly during turbulent periods (Pastor and Veronesi, 2012; Jurado et al., 2015). These findings motivate our treatment of VIX as a state variable to condition the forecasts of ERP predictors.

We employ VIX to classify each month in the out-of-sample period into two categories: months with normal levels of economic uncertainty and months with abnormally high levels of economic uncertainty.[3] Then, we evaluate the forecast accuracy of individual predictors in two different regimes of economic uncertainty (normal/high), thereby conditioning the forecasts on economic uncertainty regimes. Thus, we obtain two predictors, one with the highest forecast accuracy in a normal economic uncertainty regime and one with the highest forecast accuracy in the high economic uncertainty regime. This optimal combination of predictors allows us to employ a switching algorithm that switches between the two predictors depending on the level of economic uncertainty.

Our dataset consists of continuously compounded monthly returns on the S&P 500 index (inclusive of dividends) and data on the standard set of predictors used in the ERP predictability literature. The sample period is from 1990 to 2020, which we partition into three sub-periods: (i) an initial estimation period (January 1990 - April 1996), (ii) a holdout sample period (May 1996 - December 2004), and (iii) a "true" out-of-sample period (January 2005 - Decmeber 2020), as shown in Figure 1.

Our prediction framework can be summarized in the following steps. First, using the standard set of predictors employed in the extant literature on ERP predictability, we begin with an initial estimation period (February 1990 - April 1996), which is recursively expanded to generate one-month ahead out-of-sample forecasts of ERP for each predictor over the period from May 1996 to December 2020. We also develop a nonparametric regime-switching model that relies on the empirical distribution of VIX to classify each out-of-sample month as belonging to either one of two regimes: (i) a normal economic uncertainty regime; (ii) a high economic uncertainty regime. Thus, for each out-of-sample month, our estimation procedure yields an ERP

[3]Used globally as a measure of economic uncertainty, VIX reflects changes in the riskiness of the stock market as well as the risk aversion (preferences) of investors. According to the capital asset pricing model (CAPM), investors expect higher returns on assets with higher systematic risk. Since the standard deviation of returns captures the systematic risk of a well-diversified portfolio of assets (e.g., the S&P 500), VIX directly influences the first moment (expected return) of the return-generating process of ERP.

forecast (for each predictor) and an economic uncertainty regime identifier of that month.

Next, we partition the out-of-sample months into a holdout sample period (May 1996 to December 2004) and a model evaluation period (January 2005 to December 2020). The holdout sample period – judiciously chosen to cover one business cycle – is used to identify the optimal predictor (i.e., the one that provides the highest forecast accuracy) in each of the two economic uncertainty regimes (normal/high). This step yields an optimal combination of two predictors, *OPT_COMB*, that has the highest forecast accuracy relative to all other alternative predictor combinations over the holdout sample period.

Finally, we evaluate the forecast accuracy of *OPT_COMB* relative to all other predictors (including the historical average ERP and the mean combination forecast) over the remaining out-of-sample period from January 2005 to December 2020.

We find that the combination of predictors that generates the highest forecast accuracy over the holdout sample period is: (i) Earnings-price ratio when the predicted one-month-ahead economic uncertainty regime is normal; (ii) Inflation when the predicted one-month-ahead economic uncertainty regime is high. We define this set of predictors as the optimal combination of predictors (*OPT_COMB*) that generates forecasts by switching between predictors depending on the economic uncertainty regime.

Our key finding is that *OPT_COMB* delivers the highest (and statistically significant) forecast accuracy relative to the historical average ERP over the out-of-sample period from 2005 to 2020. Neither the mean combination forecast nor the individual predictors outperforms the historical average ERP over the same period. Therefore, we find that the forecast accuracy of individual predictors varies across economic conditions, highlighting how investors could prioritize predictors depending on the level of economic uncertainty.

We provide novel evidence that the predictive power of the earnings-price ratio and inflation for ERP is regime-dependent and varies systematically with macroeconomic uncertainty. Specifically, we show that the earnings-price ratio forecasts ERP out-of-sample during periods of normal economic uncertainty, while inflation becomes the dominant predictor during high-uncertainty regimes. This asymmetry aligns with theoretical models that link expected returns to time-varying risk aversion and macroeconomic volatility (Campbell and Cochrane, 1999; Bansal and Yaron, 2004; Piazzesi and Schneider, 2007; Campbell et al., 2018), although it has not been previously documented in the empirical literature. While prior studies highlight the conditional performance of individual predictors (Bekaert and Engstrom, 2010; Henkel et al., 2011; Dangl and Halling, 2012; Pettenuzzo et al., 2014; Castelnuovo et al., 2023), our contribution is to show that different predictors dominate in different uncertainty regimes and that this structure significantly enhances out-of-sample forecast accuracy. Therefore, our findings underscore the importance of accounting for economic uncertainty when designing predictive models and interpreting valuation ratios and macroeconomic indicators in asset pricing.

The state-dependent predictive accuracy of individual predictors aligns with consumption-based asset pricing models, in which the marginal utility of consumption shapes the pricing of risk. The earnings-price ratio predicts returns during periods of normal uncertainty when the marginal utility of consumption is relatively low, consistent with models in which expected returns are closely linked to fundamentals and risk premia are stable (Campbell and Cochrane, 1999; Bansal and Yaron, 2004). In contrast, during periods of elevated uncertainty – when macroeconomic risks and the marginal utility of consumption rise – inflation becomes the dominant predictor, reflecting investor concerns about real return erosion and time-varying risk aversion (Piazzesi and Schneider, 2007; Campbell et al., 2018). For investors, our results highlight the value of adapting predictive strategies to prevailing macroeconomic conditions to better assess compensation for risk.

Our findings also have important implications for policymakers. During periods of normal economic uncertainty, monetary policy can influence the equity risk premium indirectly through its impact on earnings growth and discount rates. However, during periods of high economic uncertainty, inflation predicts returns in line with the inflation bias mechanism of Cukierman and Gerlach (2003), where precautionary monetary policy under uncertainty leads to systematically higher inflation and an increased inflation risk premium. In particular, the inflation bias mechanism implies that policy responses under high uncertainty can increase asset price sensitivity to monetary shocks (Bernanke and Kuttner, 2005; Bekaert et al., 2013). Therefore, during periods characterized by high economic uncertainty, monetary interventions that neglect the inflation-risk premium channel may inadvertently amplify asset price volatility, complicating policy transmission without improving macroeconomic stabilization.

We also demonstrate the economic significance of our findings by computing utility gains (relative to the historical average ERP) and Sharpe ratios generated by the predictors when their forecasts are utilized through a trading strategy. We find that the utility gain realized from *OPT_COMB* is larger than those realized from the individual predictors and the mean combination forecast across all reasonable values of investor risk aversion. In particular, *OPT_COMB* generates utility gains of around 4%. Using the haircut Sharpe ratio – the Sharpe ratio adjusted for data mining biases – as an alternative measure of risk-adjusted returns, we find that *OPT_COMB* generates the highest haircut Sharpe Ratio of 0.41.

In our final analysis, we examine whether the key results hold when we employ alternative approaches to classify the economic uncertainty regimes over the period for which we generate ERP forecasts. Overall, the results from these tests show that *OPT_COMB* continues to deliver the highest forecast accuracy, regardless of the regime-classification approach. Consequently, we are able to propose a robust forecasting strategy that is useful to investors in real time.

We contribute to the existing literature on ERP predictability in four ways. First, we show that the prediction accuracy of individual predictors depends on the economic uncertainty environment, and, therefore, it may *not* be optimal to predict ERP using a single predictor throughout the out-of-sample period. Second, the algorithm we propose to identify the optimal combination of predictors is tractable and generalizable to not only other developed markets but also emerging markets where economic uncertainty is likely to be more prescient. Third, as highlighted by Wolff and Neugebauer (2019) and Denk and Löffler (2024), even complex machine learning models for equity premium prediction fail to outperform the historical average benchmark over an out-of-sample period that largely coincides with our sample period. In contrast, our economic-intuition based model is simple to implement and performs better than other relatively complex models, which makes our model particularly useful for practitioners.[4] Finally, we analyze one-month-ahead ERP forecasts, whereas most prior studies (Goyal and Welch, 2003; Welch and Goyal, 2008; Rapach et al., 2010) evaluate quarterly or annual forecasts. Our findings are therefore relevant for industry stakeholders who require estimates of the equity risk premium at a higher frequency, such as portfolio managers who rebalance their portfolios frequently and corporate managers who assess short-term investments.

The rest of the paper is organized as follows. In Section 1, we present the econometric methodology that is employed to generate one-month-ahead ERP forecasts for each predictor from May 1996 to December 2020. We also benchmark our findings to the existing literature and show that our results are consistent with the findings of Welch and Goyal (2008), Rapach et al. (2010), and Denk and Löffler (2024). In Section 2, we classify out-of-sample months as belonging to a normal economic uncertainty regime or a high economic uncertainty regime. In Section 3, we discuss how we identify the

[4]An exception is Gu et al. (2020), who show that regression trees and neural networks yield forecasting improvements comparable to those achieved by our optimal predictor combination. However, the out-of-sample period considered in their study (1987 - 2016) overlaps by approximately 40% with the sample period used in our analysis, limiting direct comparability. Moreover, machine learning methods raise concerns regarding interpretability, as they typically do not reveal the underlying economic mechanisms linking asset prices to conditioning variables.

optimal combination of predictors (*OPT_COMB*) over the holdout sample period from May 1996 to December 2004. In Section 4, we discuss the performance of *OPT_COMB* over the true out-of-sample period from January 2005 to December 2020. Finally, in Section 5, we demonstrate the economic significance of forecasting gains from using *OPT_COMB*. Section 6 reports the results of the robustness tests, and Section 7 concludes.

1 Generation of ERP Forecasts and Forecast Evaluation

To predict the **equity risk premium (*ERP*)** at a monthly frequency, we consider the 14 individual predictors used in Welch and Goyal (2008), for which monthly data are available from January 1990 to December 2020.[5] This dataset is used in the paper of Welch and Goyal (2008) and a majority of subsequent papers on ERP predictability. Our start period is January 1990 because information on VIX (the proxy for economic uncertainty) is available from 1990 onward. Moreover, we use monthly rather than quarterly or annual data to: (i) maximize the sample size for forecast evaluation; (ii) render our work more useful for practitioners interested in higher frequency estimates of ERP.

Our predicted variable is ERP, which is the total rate of return on the stock market index (the S&P 500 index) minus the risk-free rate. Stock returns are measured as the continuously compounded monthly returns on the S&P 500 index (inclusive of dividends), and the risk-free rate is the 3-month Treasury bill rate. The predictor variables are related to stock characteristics, interest rates, and the macroeconomic environment.[6] Table

[5] Welch and Goyal (2008) provide detailed descriptions of the data and their sources. The updated dataset (downloaded in September 2021) is available at `https://sites.google.com/view/agoyal145`.

[6] One indicator of the macroeconomic environment that is employed by Welch and Goyal (2008) to predict equity risk premium is the investment-to-capital ratio, measured as the ratio of private non-residential fixed investment to aggregate capital for the overall

1 reports the variables and their definitions. Table 2 reports the summary statistics.

We investigate the predictability of ERP at a monthly horizon using a standard linear regression model, which is expressed as:

$$r_{t+1} = \alpha_i + \beta_i x_{i,t} + \eta_{t+1} \tag{1}$$

where r_{t+1} is the monthly ERP, $x_{i,t}$ is a predictor whose prediction accuracy is investigated, and η_{t+1} is assumed to follow a white noise process.[7] Following Welch and Goyal (2008) and Rapach et al. (2010), we obtain out-of-sample forecasts of ERP using a recursive estimation window. Starting with the first k in-sample observations for r_t and $x_{i,t}$, the ERP forecast for month $k+1$ based on predictor $x_{i,t}$ is given by:

$$\hat{r}_{i,k+1} = \hat{\alpha}_{i,k} + \hat{\beta}_{i,k} x_{i,k} \tag{2}$$

where $\hat{\alpha}_{i,k}$ and $\hat{\beta}_{i,k}$ are the ordinary least squares (OLS) estimates of α_i and β_i, respectively, in equation (1). These estimates are obtained by regressing $[r_t]_{t=2}^{k}$ on a constant and $[x_{i,t}]_{t=1}^{k-1}$. Similarly, the ERP forecast for month $k+2$ is obtained by regressing $[r_t]_{t=2}^{k+1}$ on a constant and $[x_{i,t}]_{t=1}^{k}$. Given a total sample size of T, we obtain $T-k$ out-of-sample forecasts of ERP based on predictor $x_{i,t}$.[8] It is important to note that this approach enables

economy. This ratio is, however, available only at a quarterly/annual frequency. Since our study focuses on shorter-horizon monthly forecasts, we do not include this variable.

[7]Prior studies on ERP predictability do not account for possible stochastic or deterministic trends of predictors. The reason for this is economic; non-stationary predictors (in a population sense) are never employed in predictive regressions since they cannot plausibly predict stationary excess returns. It is reasonable to assume that the predictors used in our analysis are stationary in the population, although some of them might appear non-stationary in a given data sample. Therefore, we do not make any adjustments for non-stationarity in the predictive regressions. We thank Amit Goyal for this valuable insight.

[8]Campbell and Thompson (2008) show that imposing restrictions on the signs of regression coefficients and return forecasts can improve the out-of-sample performance of individual predictive regression models. However, Rapach et al. (2010) find that these restrictions have relatively little effect on the forecast accuracy of individual predictors

an investor to predict the equity risk premium in real time.

In our analysis, we obtain monthly out-of-sample forecasts of ERP based on each of the 14 predictors ($i = 1, 2, ..., 14$), which are outlined in Table 1. Moreover, we start with an initial estimation period of 75 months ($k = 76$) in the baseline regressions and recursively obtain 296 one-month-ahead out-of-sample forecasts of ERP for each predictor from May 1996 to December 2020.

Welch and Goyal (2008) show that the historical average ERP has superior forecast accuracy compared to the aforementioned 14 predictors over the out-of-sample period from 1965 to 2004. Building on their work, Rapach et al. (2010) find that the mean combination forecast outperforms both the historical average ERP and the 14 individual predictors over the same out-of-sample period. Therefore, we consider these two benchmark predictors in our analysis, which are computed as follows:

Historical Average Equity Risk Premium (*HIST_MEAN*): The average of in-sample observations for the equity risk premium in each recursive estimation window (the historical average). This average is the one-month-ahead ERP forecast and serves as the first benchmark predictor.

Mean Combination Forecast (*MEAN_COMB*): Following Rapach et al. (2010), we compute *MEAN_COMB* as the equally weighted average of the ERP forecasts of the 14 predictors.[9] Rapach et al. (2010) argue that the mean combination forecast generates superior forecast accuracy relative to the historical average ERP and individual predictors because it captures different aspects of the complex return-generating process of ERP by combining information obtained from multiple predictors. At the same time, the

and combination forecasts. Similarly, Denk and Löffler (2024) do not find evidence that these restrictions improve the predictive ability of combination forecasts in recent decades. Therefore, we do not impose the Campbell-Thompson restrictions in our predictive models.

[9] Rapach et al. (2010) evaluate several alternative combination forecasts such as the trimmed mean combination forecast, the median combination forecast, and the discounted mean square prediction error (DMSPE) forecast. They find that the simplest method for combining forecasts – the mean combination – consistently outperforms the more complicated methods. Therefore, among these alternatives, we choose the mean combination forecast as the second benchmark predictor.

volatility of combination forecasts is reduced due to diversification benefits. Moreover, the mean combination forecast is closely tied to business cycles, which explains the economic sources of predictability.

We compare the forecast accuracy of the predictors relative to the historical average ERP (*HIST_MEAN*) by using the R^2_{OS} statistic, proposed by Campbell and Thompson (2008). Denoting $\hat{r}_t$ and $\bar{r}_t$ as the ERP forecasts of a predictor and *HIST_MEAN*, respectively, the R^2_{OS} statistic is given by:

$$R^2_{OS} = 1 - \frac{\sum_{t=1}^{T-k}(r_{k+t} - \hat{r}_{k+t})^2}{\sum_{t=1}^{T-k}(r_{k+t} - \bar{r}_{k+t})^2} \tag{3}$$

where $T - k$ is the number of one-month-ahead out-of-sample forecasts. A positive value of R^2_{OS} implies that the predictor has lower mean squared predictor error (MSPE) than *HIST_MEAN*. We test whether this lower MSPE is statistically significant, which is equivalent to testing the null hypothesis that $R^2_{OS} \leq 0$ against the alternative hypothesis that $R^2_{OS} > 0$. The standard approach for testing this hypothesis is to use the Diebold and Mariano (1995) statistic. However, the statistic is correctly sized only for non-nested models and is severely undersized for nested models (Diebold and Mariano, 2002; Clark and West, 2007). Note that our predictive regression model is nested, as the ERP forecasts of the predictors reduce to the historical average ERP when we restrict $\beta_i = 0$ in equation (1). Therefore, we rely on the methodology of Clark and West (2007), who propose an adjusted version of the Diebold and Mariano (1995) statistic - the MSPE-adjusted statistic - that is correctly sized when comparing forecasts from nested models. The MSPE-adjusted statistic is computed as:

$$MSPE{-}adjusted = \frac{1}{T-k}\sum_{t=1}^{T-k}\{(r_{k+t}-\bar{r}_{k+t})^2-(r_{k+t}-\hat{r}_{k+t})^2+(\bar{r}_{k+t}-\hat{r}_{k+t})^2\} \tag{4}$$

Using the MSPE-adjusted statistic, we test whether the population MSPEs of two models are equal, which is akin to testing whether $\beta_1 = 0$ in equation

(1). The statistic is computed as the difference between sample MSPEs plus an adjustment term which removes the noise in ERP forecasts introduced by estimating the additional parameter, β, when it is in fact zero in the population model. The adjustment term is the average of the squared difference between forecasts based on $HIST\_MEAN$ and the predictor, which is shown as the third term in equation (4). Clark and West (2007) show that the MSPE-adjusted statistic has an asymptotically standard normal distribution. For finite sample sizes, however, the statistic has a nonstandard distribution.

Given that our sample is moderately sized, we estimate more precise p-values by the nonparametric bootstrap procedure.[10] Specifically, we construct the sampling distribution of the MSPE-adjusted statistic by sampling randomly with replacement 10,000 times from the demeaned out-of-sample forecasts, each time computing the MSPE-adjusted statistic.[11] We then calculate the area in the right tail (one-sided p-value) of the bootstrapped sampling distribution corresponding to the original sample statistic.

Table 3 reports R^2_{OS} statistics for the 14 individual predictors and the mean combination forecast ($MEAN\_COMB$) over the out-of-sample period from May 1996 to December 2020. We also report the corresponding p-values obtained from the standard normal and the nonparametric bootstrap distributions of the MSPE-adjusted statistic.

Panel A of Table 3 shows that none of the 14 individual predictors are able to outperform the historical average ERP, as evidenced by the negative R^2_{OS} statistics (ranging from -6.12% to -0.17%). This is consistent with the findings of Welch and Goyal (2008) and Rapach et al. (2010), who document that the predictive ability of individual predictors deteriorates considerably over the 1976-2005 out-of-sample period. Indeed, Rapach et al. (2010) find that the R^2_{OS} statistic is marginally positive (but not statistically significant)

[10] The nonparametric bootstrap procedure assumes that out-of-sample forecasts are independent. The assumption holds in our context since optimal one-month-ahead forecasts are independent (Diebold and Mariano, 1995; Clark and West, 2007).

[11] Demeaning ensures that the sampling distribution of the statistic is centered at zero.

for only one of the individual predictors (*NTIS*) over this period. They attribute this decline in forecast accuracy to multiple structural breaks in the predictive regression models occurring in the mid-1970s, corresponding to the Oil Shocks, and the mid-1980s, after the change in Federal Reserve operating procedures. Although the out-of-sample period in our analysis (1996M5 - 2020M12) is mostly non-overlapping with that in Rapach et al. (2010) (1976Q1 - 2005Q4), these arguments also apply to our context owing to multiple structural breaks observed during the dot-com bubble (late-1990s to early 2000), the dot-com bust (2000-01), the boom leading up to the Global Financial Crisis (GFC), the GFC itself (2007-09), the extended period of monetary easing in advanced economies since the GFC, interrupted by the "taper tantrum" episode in 2013, and the early stages of the COVID-19 pandemic.[12]

Panel B of Table 3 shows that the mean combination forecast is also unable to outperform the historical average ERP, as evidenced by the negative R^2_{OS} statistic of -0.52%. This is consistent with the findings of Denk and Löffler (2024), who find that combination forecasts – proposed by Rapach et al. (2010) – fail to outperform the historical average ERP in the recent period from 1994 to 2022. Specifically, for the mean combination forecast, Denk and Löffler (2024) also find a negative R^2_{OS} statistic of -0.27%. They attribute the decline in predictive ability to an increase in the correlation of forecast errors of individual predictive regression models, thereby reducing the potential of combination forecasts to reduce forecast volatility through diversification. Moreover, the increased correlations imply that the individual predictors capture similar information on the return-generating process of ERP, so combining forecasts of individual predictors does not enhance the information content. Importantly, their analysis does not indicate that the decline in predictive accuracy is due to an unfavorable forecast environment. In general, the lack of robustness of the mean combination forecast confirms

[12]Note that we generate monthly forecasts while Rapach et al. (2010) generate quarterly forecasts. Therefore, our results may not be strictly comparable.

the broader conclusion in the literature that the findings from studies published years ago change when more recent data is used (Welch and Goyal, 2008).

2 Classification of Out-of-sample Months as Normal/High VIX Periods

Predicting the dynamic return-generating process of ERP by employing a *single* predictor throughout the out-of-sample period is not always the ideal forecasting approach, as indicated by the analysis in Section 1. Therefore, in order to obtain superior forecasts, we develop a prediction framework that allows us to optimally switch between predictors based on key parameters that could affect the return-generating process of ERP.

We argue that economic uncertainty is one of the key parameters driving changes in the return-generating process of ERP. We measure economic uncertainty by the Volatility Index (VIX), which is a widely used indicator of economic uncertainty that reflects changes in the riskiness of the stock market as well as the risk aversion (preferences) of investors. VIX directly influences the first moment (expected return) of the return-generating process of ERP due to the positive association between risk aversion and equity risk premium (Fama and French, 1989; Cochrane, 1999, 2007). However, it is seldom used as a direct predictor of ERP, due to its limited forecasting power and nonlinear relationship with future returns. As Bollerslev et al. (2009) show, while option-implied volatility reflects valuable information about variance risk premia, it does not reliably forecast excess returns in linear models.

Instead, VIX serves more effectively as a regime classification tool, capturing high-frequency shifts in market sentiment, risk aversion, and macroeconomic uncertainty that correspond to structural changes in the pricing of risk. Bekaert and Hoerova (2014) demonstrate that increases in VIX are strongly associated with deteriorating macro-financial conditions and that the variance risk premium component of VIX – distinct from the conditional stock market variance component – contains predictive content for investors' time-varying compensation for risk. Moreover, recent work highlights the threshold-dependent nature of VIX's information content. For example, Bansal and Stivers (2025) find that ERP increases meaningfully only when VIX exceeds its 80^{th} percentile, suggesting VIX's role as a nonlinear trigger for regime changes rather than as a continuous predictor. Additionally, Dennis et al. (2006) document that negative shocks to VIX are followed by higher ERP, reinforcing the asymmetric and nonlinear nature of its relation to returns. Pastor and Veronesi (2012) and Jurado et al. (2015) further argue that uncertainty shocks – proxied in part by VIX – are central drivers of asset price dynamics, particularly during turbulent periods. These insights motivate our treatment of VIX as a state variable, allowing us to condition the performance of ERP predictors on the prevailing level of economic uncertainty.

Panel A of Figure 2 plots the time series of VIX observations from January 1990 to December 2020. The shaded vertical bands denote economic recessions as estimated by the National Bureau of Economic Research (NBER). We observe that VIX captures changes in economic uncertainty quite well, as evidenced by the noticeable spikes during NBER-dated recessions, such as the GFC and the COVID-19 pandemic (March 2020 onward). Furthermore, Panel B shows that the VIX time series is highly persistent (1^{st} and 2^{nd} order autocorrelation of 0.9 and 0.8, respectively). This implies that abrupt changes in VIX are rare, which makes it suitable as an economic uncertainty regime indicator.

Building on our argument, we develop a nonparametric regime-switching model to classify the economic uncertainty regime for each month from May 1996 to December 2020 (the out-of-sample period). Specifically, we classify VIX observations in the out-of-sample period as belonging to a normal or high economic uncertainty regime. To do this, we compute the quantile of VIX for each month t (VIX_t) by using the empirical distribution of VIX observations in the prior 75 months; a rolling window where the window size is chosen to match the initial estimation period. If the quantile of VIX_t is less (more) than the 75th percentile of the corresponding distribution of in-sample VIX observations, we classify the economic uncertainty regime for out-of-sample month $t + 1$ as normal (high). Since VIX_t is an estimate of the volatility of stock returns in month $t + 1$, we are, in effect, predicting the one-month-ahead economic uncertainty regime. Moreover, t runs from April 1996 up to November 2020, which enables us to classify economic uncertainty regimes for each month in the out-of-sample period.

An example might be useful to illustrate our regime-switching model. Suppose that we are in April 1996 and wish to classify the economic uncertainty regime for May 1996 (the first out-of-sample month). Following the approach mentioned earlier, we compute the quantile of VIX for April 1996, which is observable, by using the empirical distribution of VIX observations from January 1990 to March 1996 (i.e., the prior 75 months). If the quantile of VIX for April 1996 is less (more) than the 75th percentile of the empirical distribution, we classify the economic uncertainty regime for May 1996 as normal (high). Similarly, we classify the economic uncertainty regime for June 1996 by using VIX for May 1996 and the empirical distribution of VIX observations from February 1996 to April 1996 (rolling windows of 75 months). Proceeding in this manner until the end of the out-of-sample period, we classify 189 months as belonging to the normal economic uncertainty regime and 107 months as belonging to the high economic uncertainty regime.

After classifying the economic uncertainty regime for each month in the out-of-sample period, we estimate the transition probability matrix of regime switches to validate our approach. We find that the probability of staying in the same regime is significantly higher than switching between regimes. More specifically, if the current month belongs to the normal uncertainty regime, then the probability that the next month also belongs to the normal uncertainty regime is 87.83%. Similarly, if the current month belongs to the high uncertainty regime, then the probability that the next month also belongs to the high uncertainty regime is 79.25%. The transition probabilities validate our regime-switching model because the predicted regimes are persistent.

Our regime-classification approach stands out in several key aspects. First, it is consistent with the approach proposed by Forbes and Warnock (2012) to identify extreme capital flow episodes (thereby capturing regime changes) in advanced and emerging economies. Second, by classifying regimes using rolling windows of VIX observations, we can capture investors' expectations of future economic uncertainty more precisely (as opposed to recursive windows). This is because investors are more likely to form expectations based on their recent experience; for example, immediately after a period of heightened volatility, such as the GFC, investors are less likely to view large but declining VIX observations as belonging to the high economic uncertainty regime when, in fact, investors have already begun incorporating the information that economic uncertainty is decreasing. Furthermore, the classification of regimes in our approach is based on monthly data, which enables us to capture periods of heightened economic uncertainty that may be missed when using quarterly or annual data. Finally, our approach is computationally inexpensive due to its nonparametric nature and, therefore, easier for investors to implement in real time.

An alternative regime-classification approach could be to model the VIX time series as a Self-Exciting Threshold Autoregressive (SETAR) process with one threshold (two regimes). The benefit of this approach is that the economic uncertainty regimes are classified using endogenously determined VIX thresholds. However, the SETAR model is data-intensive, implying that we are unlikely to obtain reliable estimates of thresholds by using rolling in-sample estimation periods of only 75 observations. Therefore, we prefer the nonparametric regime-switching model.

3 Identification of the Optimal Combination of Predictors

In this section, we describe the approach to identify the predictors that demonstrate the best forecast accuracy in each of the two regimes of economic uncertainty (normal and high). Put differently, we condition the forecasts of the predictors on economic uncertainty in order to identify the best predictor in each of the two regimes. This step follows from the intuition that, given the economic uncertainty regime, certain predictors may outperform all other predictors in that regime. For instance, forecasts based on TBL and DY in the normal and high economic uncertainty regimes, respectively, may be more accurate than forecasts based on any single predictor (including the mean combination forecast and the historical average ERP) over the entire out-of-sample period. However, we cannot identify, ex-ante, the most accurate predictor in each of the two economic uncertainty regimes. To circumvent this issue, we select the first 104 months (May 1996 - December 2004) – from the period for which we generate one-month-ahead ERP forecasts – as the holdout sample period. We then identify the most accurate predictor in each of the two economic uncertainty regimes over the holdout sample period. This 2-tuple predictor combination is defined as the optimal combination of predictors (OPT_COMB). We use the remaining 192 months (January 2005 - December 2020) as the model evaluation period (henceforth, the out-of-sample period).

The procedure for identifying the optimal combination implicitly assumes that the return-generating process of ERP changes across the two economic uncertainty regimes, which implies that it may not be optimal to track the dynamic return-generating process of ERP by employing a single predictor over the entire out-of-sample period. Furthermore, this procedure is similar in spirit to the model selection (ms) approach in Welch and Goyal (2008). Given N predictors, the model selection approach considers 2^N predictive regression models that include all possible combinations of the predictors.[13] Then, for each period t, the model that demonstrates the minimum cumulative out-of-sample forecast error up to time t is chosen to generate out-of-sample forecasts.

We choose the holdout sample period (particularly the end date) to correspond to an NBER-dated business cycle, as shown in Figure 3. This choice is motivated by economic and statistical considerations. Since business cycle upturns (downturns) are generally associated with a decrease (increase) in economic uncertainty (Fama and French, 1989; Cochrane, 1999, 2007), choosing the holdout sample period that coincides with a business cycle allows us to capture enough months of normal and high economic uncertainty. This is important because the forecast evaluation tests are more powerful if we have a reasonable number of observations in each of the two economic uncertainty regimes. An additional benefit of restricting the holdout sample period to one business cycle is that we are left with a sufficient number of observations (January 2005 - December 2020) to evaluate out-of-sample forecasts reliably.

Table 4 reports R^2_{OS} statistics for the predictors in the two economic uncertainty regimes over the holdout sample period from May 1996 to December 2004. We draw statistical inferences by referring to the p-values obtained from the nonparametric bootstrap distribution of the MSPE-adjusted statistic. Panels A and B of Table 4 report the results for the 14 individual predictors and *MEAN_COMB*, respectively. We find that *EP* delivers the highest forecast accuracy (largest R^2_{OS} statistic) during periods of normal economic uncertainty, while *INFL* delivers the highest forecast accuracy during periods of high economic uncertainty. Moreover, the R^2_{OS} statistic of *EP* in the normal uncertainty regime is sizeable (14.87%) and strongly statistically significant (at the 1% level). The R^2_{OS} statistic of *INFL* is moderately large (0.16%) but statistically insignificant in the high uncertainty regime. Since our statistical tests are expected to have poor power in the high un-

[13] The predictive models include both simple and multiple linear regression models.

certainty regime – which comprises tail outcomes with fewer observations by definition – this is not especially problematic.

Our results for the holdout sample period indicate that the highest forecast accuracy is generated by: (i) the earnings-price ratio (*EP*) when the predicted one-month-ahead economic uncertainty regime is normal; (ii) inflation (*INFL*) when the predicted one-month-ahead economic uncertainty regime is high. Accordingly, we define the 2-tuple predictor combination [*EP*, *INFL*] as the optimal combination of predictors (*OPT_COMB*). Since *OPT_COMB* delivers the best forecast accuracy over the holdout sample period spanning one business cycle, we expect it to perform similarly over the out-of-sample period from 2005 to 2020 that includes multiple business cycles.

For robustness, we also consider a longer holdout sample period that covers two NBER-dated business cycles from May 1996 to December 2014. Appendix Table A.1 reports R^2_{OS} statistics for the predictors in the two economic uncertainty regimes over the longer holdout sample period. We continue to find that *EP* delivers the highest forecast accuracy in the normal economic uncertainty regime, while *INFL* delivers the highest forecast accuracy in the high economic uncertainty regime. This shows that the optimal combination of predictors is not sensitive to our choice of the holdout sample period. Moreover, the superior forecast accuracy of both *EP* and *INFL* is strongly statistically significant, possibly because forecast evaluation tests are more powerful when sample sizes are larger. However, a major drawback of the longer holdout sample period is that it leaves us with only 72 months (January 2015 to December 2020) to conduct reliable out-of-sample tests. For this reason, we prefer the holdout sample period that spans one business cycle so that there is a sufficient number of observations in the out-of-sample period. Of course, a longer history of observations offers the flexibility of choosing a holdout sample period that covers multiple business cycles.

4 Forecast Accuracy of *OPT_COMB*

In this section, we evaluate the forecast accuracy of *OPT_COMB*, relative to the individual and benchmark predictors, over the out-of-sample period from January 2005 to December 2020. Since we have already classified the economic uncertainty regimes over the out-of-sample period, we simply choose the ERP forecast of the best predictor corresponding to the regime for the out-of-sample month, as identified by *OPT_COMB*. This generates out-of-sample forecasts for *OPT_COMB*, which enables us to evaluate its forecast accuracy.

Table 5 reports R^2_{OS} statistics for the 14 individual predictors, the mean combination forecast, and *OPT_COMB* over the out-of-sample period from January 2005 to December 2020. We also report the corresponding p-values obtained from the standard normal and the nonparametric bootstrap distributions of the MSPE-adjusted statistic. However, we discuss the results by referring to the more precise p-values obtained from the latter distribution.[14]

Panel A of Table 5 shows that the R^2_{OS} statistics of the 14 individual predictors range from -8.87% to 0.49%, with only three among them (DP, DY, and BM) having a positive value. However, none of the positive R^2_{OS} statistics are statistically significant at conventional levels. Although the predictive ability of a majority of individual predictors improves marginally over the more recent out-of-sample period relative to their performance over the out-of-sample period from May 1996 to December 2020 (Panel A of Table 3), they still fail to outperform the historical average ERP. Panel B of Table 5 shows that the mean combination forecast is also unable to outperform the historical average ERP, as evidenced by the negative R^2_{OS} statistic of -0.60%. Furthermore, the predictive ability of the mean combination forecast worsens over the more recent out-of-sample period compared to its performance over the out-of-sample period from May 1996 to December 2020 (Panel B of Table

[14]Our key findings do not change when we refer to the p-values from the standard normal distribution since they are similar to those computed from the nonparametric bootstrap distribution.

3). The reasons for the underperformance of individual predictors and the mean combination forecast are similar to those discussed in Section 1.

Turning to the results for *OPT_COMB* in Panel C of Table 5, we find that the R^2_{OS} statistic is not only positive and statistically significant at the 10% level (p-value = 6.81%, which is closer to the 5% level), but it is also sizeable (1.33%). Thus, *OPT_COMB* is the only predictor that is able to reliably outperform the historical average ERP over the recent out-of-sample period. Interestingly, the forecast accuracy of *OPT_COMB* is also superior to that of the mean combination forecast – the predictor shown to outperform all individual predictors and the historical average ERP by Rapach et al. (2010). The superior forecast accuracy is easy to verify given the opposing signs and statistical significance (or lack thereof) of the R^2_{OS} statistic for the two predictors.

It is instructive to benchmark our results for *OPT_COMB* with those in Gu et al. (2020) and Denk and Löffler (2024). Gu et al. (2020) explore a plethora of machine learning methods and find that regression trees (including boosted trees and random forests) and neural networks yield monthly R^2_{OS} statistics in the range of 1.08% to 1.80%, which are very similar to that generated by *OPT_COMB*. However, these results are not strictly comparable because the out-of-sample period considered in their study (1987 - 2016) overlaps by approximately 40% with the sample period used in our analysis. More importantly, machine learning methods often lack interpretability and do not reveal the economic relationships between asset returns and predictor variables. Similarly, in order to revive the utility of combination forecasts, Denk and Löffler (2024) try several forecasting methods from the machine learning literature and multiple variations based on data frequency, estimation sample, predictor set, and forecast horizon. Among multiple forecasting methods, they employ kitchen sink regressions (a model in which all individual predictors are included), iterated combination forecasts, penalized regressions, and dimension reduction approaches. However, none of these methods leads to significant accuracy gains relative to the historical average

ERP after 1994. Considering their findings, we make an important contribution to the literature by proposing an economic-intuition based model that is simple to implement yet effective.

In Figure 4, we illustrate the forecasting gains of *OPT_COMB* relative to the benchmark predictors – the historical average ERP and the mean combination forecast – by plotting the time series of the differences between cumulative squared forecast error for the benchmark predictors and the cumulative squared forecast error for *OPT_COMB*. These plots provide a visual impression of the consistency of a predictor's out-of-sample forecasting performance. When the curve in each panel of Figure 4 trends upwards, the forecasts based on *OPT_COMB* outperforms the historical average ERP and *MEAN_COMB*, while the opposite holds when the curve trends downwards. The plots also allow us to compare the relative forecast accuracy of *OPT_COMB* for any particular out-of-sample period by redrawing the horizontal zero line to correspond to the start of the out-of-sample period. Essentially, we compare the height of the curve at the two points corresponding to the beginning and end of a given out-of-sample period: if the curve is higher (lower) at the end of the out-of-sample period than at the beginning, the forecasts based on *OPT_COMB* has a lower (higher) MSPE over the out-of-sample period. A predictor that always outperforms a chosen benchmark predictor for any out-of-sample period will have a positively-sloped curve throughout; this serves as a useful reference to evaluate the ability of a predictor to consistently beat the benchmark predictor.

Panel A of Figure 4 plots the difference between the cumulative squared forecast error for the historical average ERP and the cumulative squared forecast error for *OPT_COMB*. The curve is predominantly positively sloped, indicating that the forecasts based on *OPT_COMB* generate out-of-sample gains relative to the historical average ERP on a consistent basis over time. Moreover, the curve is often strongly positively sloped from mid 2008 to 2009, more moderately but still consistently positively sloped from 2010 to 2018, and slightly negatively sloped from 2005 to early 2008 and during the

COVID-19 crisis in 2020. Panel B of Figure 4 plots the difference between the cumulative squared forecast error for the mean combination forecast and the cumulative squared forecast error for *OPT_COMB*. Generally, the curve follows a similar pattern as the curve in Panel A, indicating that the forecasts based on *OPT_COMB* also generate out-of-sample gains relative to the mean combination forecast on a consistent basis over time. The only notable difference is that the curve is negatively sloped for a brief period during the GFC in early 2009.

A drawback of Figure 4 is that it does not readily provide information on the magnitude and the duration of outperformance of *OPT_COMB* relative to the benchmark predictors. More generally, a predictor may outperform the benchmark predictor by a small margin for most of the out-of-sample period but underperform by a wide margin for the remaining period.[15] The forecasts for an ideal predictor should outperform the benchmark predictor by a wide margin throughout the out-of-sample period. In Figure 5, we illustrate how close *OPT_COMB* is to the ideal predictor by plotting the histogram of the difference between squared forecast error for the benchmark predictors (i.e., historical average ERP and the mean combination forecast) and the squared forecast error for *OPT_COMB*.

Panel A of Figure 5 plots the histogram of the difference in squared forecast error for the historical average ERP and the squared forecast error for *OPT_COMB*. We observe that the distribution is clustered around zero and has a positive mode, indicating that *OPT_COMB* outperforms the historical average ERP by a small margin for a majority of months in the out-of-sample period. When it underperforms, it does so by a small margin. Moreover, *OPT_COMB* outperforms the historical average ERP by a wide margin for some months in the out-of-sample period, as evidenced by the positively skewed distribution. Panel B of Figure 5 plots the histogram of

[15]The plot of the cumulative squared forecast error difference for such a predictor would show up as a gradual upward-sloping curve for most of the out-of-sample period with infrequent and substantial falloffs that last for short periods.

the difference in squared forecast error for the mean combination forecast and the squared forecast error for OPT_COMB. The distribution is slightly more positively skewed, but otherwise follows a similar pattern as the distribution in Panel A.

The key findings from Table 5 and Figures 4 and 5 are as follows:

- The results in Panel A of Table 5 reinforce the findings of Welch and Goyal (2008) and Rapach et al. (2010) by showing that individual predictors are unable to generate reliable out-of-sample forecasts. The results in Panel B of Table 5 challenge the main finding of Rapach et al. (2010) by demonstrating that the mean combination forecast fails to consistently generate out-of-sample forecasting gains compared to the historical average ERP over the recent period (2005 - 2020).

- The optimal combination of predictors (OPT_COMB) consistently outperforms both the historical average ERP and the mean combination forecast. The outperformance of OPT_COMB can be attributed to the fact that it tracks the dynamic return-generating process of ERP by employing *multiple* predictors.

In summary, we have developed an effective procedure for forecasting the equity risk premium that outperforms the benchmark predictors. The optimal forecasting strategy is to first classify the one-month-ahead level of economic uncertainty into normal or high uncertainty regimes and then choose the ERP forecast of the best predictor corresponding to that regime, as identified by OPT_COMB.

A growing literature emphasizes that the predictive power of valuation ratios for future equity returns is state-dependent, varying systematically with the level of macroeconomic uncertainty. Theoretically, models with long-run risks (Bansal and Yaron, 2004) and external habit formation (Campbell and Cochrane, 1999) suggest that valuation ratios such as the earnings-price ratio more accurately reflect expected returns during periods of low macroeconomic volatility – when consumption growth is smooth, risk aversion is subdued, and the surplus consumption ratio is high. Under such conditions, risk premia are more stable and valuation ratios are more tightly linked to expected returns.

Empirical research supports these predictions. Henkel et al. (2011) find that the earnings-price ratio forecasts short-horizon equity returns only in low-volatility regimes identified via a regime-switching volatility model. Dangl and Halling (2012) show that the slope on the earnings-price ratio is substantially higher during stable periods using a time-varying coefficient model, while Pettenuzzo et al. (2014) demonstrate improved forecasting performance when excluding crisis episodes. Together, these results suggest that the earnings-price ratio is most informative in stable environments characterized by normal levels of economic uncertainty.

Recent research also highlights that macroeconomic variables, particularly inflation, can exhibit state-dependent predictive power for ERP, especially during periods of heightened uncertainty. In long-run risk models, inflation captures concerns over real return erosion and reflects priced macroeconomic risks that become more salient under time-varying volatility (Bansal and Yaron, 2004; Campbell et al., 2018). Moreover, when the credibility of monetary policy is uncertain, inflation can influence asset prices through their effect on the term structure and risk premia (Piazzesi and Schneider, 2007).

Empirical evidence also supports a conditional role for inflation in predicting ERP. Bekaert and Engstrom (2010) find that elevated economic uncertainty often coincides with high inflation and increased risk aversion. Castelnuovo et al. (2023) show that the positive correlation between inflation and uncertainty holds primarily in high-inflation environments, further reinforcing the notion of state dependence. Since risk aversion is a key channel through which macroeconomic uncertainty affects expected returns (Fama and French, 1989; Cochrane, 1999, 2007), these findings suggest that inflation's predictive power is greatest when uncertainty – and thus the marginal value of information – is elevated. Even without conditioning on economic uncertainty, Brandt and Wang (2003) show that inflation exhibits a positive covariance with risk aversion and, by extension, ERP.[16] Consistent with this channel, Campbell and Vuolteenaho (2004) find that inflation explains approximately 80% of the time-series variation in stock-market mispricing.

Despite this growing literature, there is little empirical work examining whether different predictors dominate across different regimes of economic uncertainty, especially in an out-of-sample setting, which is the relevant standard for forecasting performance. This is where our paper makes a novel contribution. We show that the earnings-price ratio and inflation predict the equity risk premium out-of-sample, but in a mutually exclusive, regime-dependent manner. Specifically, the earnings-price ratio forecasts ERP during periods of normal economic uncertainty, while inflation becomes the dominant predictor during high-uncertainty regimes. This pattern is not captured in linear models that average over economic states and has not been previously documented in the literature.

By uncovering this asymmetry, our findings advance the return predictability literature along two key dimensions. First, we show that economic uncertainty acts as a regime-switching mechanism that determines the relevance of competing predictors. Second, we demonstrate that conditioning on economic uncertainty substantially enhances out-of-sample forecasting performance, with important implications for the design of predictive models and the interpretation of valuation ratios and macroeconomic variables in asset pricing.

The nonlinear, state-dependent predictive accuracy of individual predictors can be understood through the lens of consumption-based asset pricing models, where the marginal utility of consumption plays a central role in shaping expected returns. In economic environments characterized by smooth consumption growth and normal levels of economic uncertainty – when the marginal utility of consumption is relatively low – the earnings-

[16] We also find a positive correlation between ERP and inflation over the full sample period from 1990 to 2020.

price ratio exhibits strong predictive power, reflecting stable risk premia and a tighter link between valuation ratios and expected returns. Conversely, during periods of heightened uncertainty and elevated consumption risk – when the marginal utility of consumption rises – investors become more sensitive to macroeconomic risks, and inflation emerges as a more informative predictor of ERP. In such states, inflation reflects concerns over real return erosion and captures shifts in investor risk aversion. These findings highlight the importance of recognizing the state-dependent nature of return predictability, as the informational content of different predictors is closely tied to the marginal value of consumption across economic uncertainty regimes.

Our finding that the earnings-price ratio and inflation forecast the equity risk premium in distinct uncertainty regimes has important policy implications. During periods of normal uncertainty, the earnings-price ratio is a reliable signal of expected returns, consistent with models where risk premia are more stable and linked closely to fundamentals (Campbell and Cochrane, 1999; Bansal and Yaron, 2004). In such environments, monetary policy can influence valuations indirectly through its impact on earnings growth and discount rates. In contrast, during periods of high economic uncertainty, inflation becomes a more informative predictor of risk premia, as investors demand compensation for real return erosion and the effects of monetary policy become more nonlinear (Piazzesi and Schneider, 2007; Campbell et al., 2018).

The state-contingent pricing of inflation is consistent with the inflation bias mechanism of Cukierman and Gerlach (2003), where uncertainty about the future state of the economy, combined with a greater concern for avoiding negative output gaps, leads central banks to err on the side of monetary expansion. Rational agents anticipate this tendency, resulting in systematically higher inflation and an upward drift in inflation risk premia. Consequently, monetary shocks can trigger stronger asset price responses in high-uncertainty regimes, complicating policy transmission and amplifying volatility in financial markets (Bernanke and Kuttner, 2005; Bekaert et al., 2013). These dynamics suggest that during periods of high economic uncertainty, policymakers should avoid overly preemptive easing and instead weigh the inflation-risk premium implications of their actions – calibrating responses to minimize excessive volatility in financial markets without undermining macroeconomic stabilization.

5 Economic Significance of Forecasting Gains from using *OPT_COMB*

Can a predictor's superior forecast accuracy be employed in a trading strategy to generate economically meaningful risk-adjusted returns for an investor? To answer this question, we estimate the utility gains (relative to the historical average ERP) and Sharpe ratios realized by an investor who uses the ERP forecasts through a trading strategy.[17]

We follow Campbell and Thompson (2008) to compute the realized utility for a mean-variance investor who rebalances their portfolio monthly between the market portfolio (the S&P 500) and risk-free 30-day T-bills based on ERP forecasts of the predictors.[18] For computing portfolio weights, we assume that the investor has a power utility function with reasonable values of the coefficient of relative risk aversion (CRRA) parameter. Moreover, we restrict the weight on the market portfolio to lie between 0 and 1.5 each month, which prevents the investor from shorting the market portfolio or leveraging more than 50%.[19] Proceeding with these assumptions, we compute the realized utility from the predictors over the out-of-sample period.

We then measure the utility gain from a predictor as the difference be-

[17]This analysis is necessary because there is no one-to-one relationship between forecast accuracy based on R^2_{OS} and the economic benefits of using the forecast (Rapach and Zhou, 2013).

[18]We assume that portfolio rebalancing occurs at the end of each month.

[19]Since the methodology for computing realized utility is common in the ERP predictability literature, we do not discuss it here in detail. Interested readers can refer to a detailed discussion in Campbell and Thompson (2008) and Rapach et al. (2010).

tween the realized utility from its forecasts and the realized utility from the historical average ERP forecast. Intuitively, the utility gain from a predictor should be positive (negative) if it is able to generate a higher (lower) risk-adjusted return relative to the historical average ERP. We multiply the utility gain by 1200 to convert it to annualized percentage utility gain.

Table 6 reports the annualized utility gain from the predictors corresponding to CRRA values of 1, 2, and 3. Panel A of Table 6 shows that only four out of the 14 individual predictors – *DP, DY, EP* and *DFY* – yield positive utility gains across all values of the relative risk aversion parameter. Panel B shows that utility gains generated by the mean combination forecast are positive only for CRRA = 3, but this gain is just 0.41%. From Panel C, we observe that utility gains generated by *OPT_COMB* are the highest for CRRA = 1 and CRRA = 2, and a close second to the utility gains generated by *EP* for CRRA = 3. The utility gains are positive and sizeable, yielding gains in the range of 2-4%. The utility gains from *OPT_COMB* are also much larger than those generated by the mean combination forecast across all values of the relative risk aversion parameter.

We also compute the Sharpe ratios generated by the predictors when their forecasts are utilized through a trading strategy.[20] For computing the Sharpe ratios, we employ the same trading strategy as that employed earlier in our analysis of realized utility gains. Furthermore, we assume that the investor has a power utility function with CRRA equal to 2. Bao (2009) shows that the Sharpe ratio for a sample of excess returns is a biased estimator of the population Sharpe ratio, and the size of the bias depends on the underlying return-generating process of excess returns. To address this, we employ an unbiased estimator of the population Sharpe ratio – proposed by Bao (2009) – to conduct right-tailed hypothesis tests. Details of the estimation procedure and its underlying assumptions are given in Appendix B.

[20]The Sharpe ratio associated with a predictor is the ratio of the sample mean and the sample standard deviation of the monthly excess returns (in excess of the risk-free rate) generated from utilizing its forecasts through the trading strategy. This measure of risk-adjusted return is widely used by practitioners to backtest trading strategies.

We then determine the predictors that generate positive and statistically significant risk-adjusted returns over the out-of-sample period. To do this, we conduct hypothesis tests on the Sharpe ratios generated by each of the 17 predictors whose forecasts are utilized through the trading strategy.[21] However, we must account for one major pitfall associated with multiple hypothesis tests; when several trading strategies are tested individually, it is expected that, by chance, some of these strategies will turn out to be profitable. In other words, these profitable strategies are likely to be false discoveries. Moreover, at a given level of significance, the probability of false discoveries (Type I error) increases with the number of trading strategies that are being tested (Harvey and Liu, 2015). Therefore, the p-value of a single test (p^S) does not reflect the true statistical significance of the Sharpe ratio generated by a trading strategy.

To address this issue related to data mining, Harvey and Liu (2015) propose the multiple hypothesis testing framework that provides an adjustment to p^S to obtain the multiple-testing-adjusted p-value (p^M). Multiple testing reduces the statistical significance of single tests because the adjusted p-value, p^M, is larger than p^S. Although there are several approaches to compute p^M, we follow the one developed by Holm (1979) as it is designed to eliminate all false discoveries.[22] Then, for each trading strategy, we estimate the haircut Sharpe ratio which is the imputed Sharpe ratio corresponding to the multiple testing p-value p^M. Since p^M is larger than p^S, the haircut Sharpe ratio is smaller than the original Sharpe ratio.

In summary, any given strategy generates a Sharpe ratio with an associated p-value p^S. We first transform p^S of the single test into a multiple-testing-adjusted p-value p^M and then estimate the haircut Sharpe ratio corresponding to p^M. Table 7 reports the annualized Sharpe ratios ($\widehat{SR}$), p^S,

[21]The 17 predictors include the 14 individual predictors, the historical average ERP, the mean combination forecast, and *OPT_COMB*.

[22]Interested readers can refer to Harvey and Liu (2015) for a detailed discussion on the multiple hypothesis testing framework, especially in the context of evaluating the out-of-sample performance of trading strategies.

p^M, and the haircut Sharpe ratios ($\widehat{HSR}$) for the trading strategies based on ERP forecasts of the 17 predictors. None of these strategies reflect transaction costs, and as such, the Sharpe ratios are overstated and should be considered as "before-costs" Sharpe ratios.

Panel A of Table 7 reports the results for the 14 individual predictors and the historical average ERP. From the second column of Panel A, we see that the Sharpe ratios generated by the predictors are in the range of 0.49 to 0.73. Individual hypothesis tests on the Sharpe ratios reveal that they are all statistically significant at conventional levels ($p^S \leq 5\%$; reported in the third column of Panel A). The fourth and fifth columns of Panel A report the multiple-testing-adjusted p-values (p^M) and the corresponding haircut Sharpe ratios, respectively. We find that the haircut Sharpe ratios are considerably smaller (around 44-63% smaller) than the original Sharpe ratios. More importantly, after applying the multiple-testing adjustment, the haircut Sharpe ratios of only two of the 15 individual predictors – EP and DFY – continue to remain statistically significant (at the 10% level of significance). These findings illustrate that the multiple-testing adjustment greatly reduces the number of profitable trading strategies, as many of them are likely to be false discoveries. Panel B shows that the haircut Sharpe ratio of the mean combination forecast is only 0.23 and is no longer statistically significant. From Panel C, we observe that OPT_COMB delivers a Sharpe ratio of 0.76 and a haircut Sharpe ratio of 0.41, with the latter remaining statistically significant at the 10% level.[23] For comparison, Gu et al. (2020) report an annualized out-of-sample Sharpe ratio of 0.77 for a market-timing strategy on the S&P 500 based on neural network forecasts, though this estimate is not adjusted for data mining bias. These results reinforce that our economically motivated combination approach performs competitively relative to more sophisticated machine learning models.

The key findings from our analysis of utility gains and haircut Sharpe

[23]The results are qualitatively similar for CRRA = 1 and CRRA = 3.

ratios are as follows:

- *OPT_COMB* delivers the highest risk-adjusted return when its forecasts are utilized through a simple trading strategy.
- Relative to utility gains, the haircut Sharpe ratio as an alternative measure of risk-adjusted returns has two important benefits: (i) we can conduct hypothesis tests for forecasting gains; (ii) we can adjust for data mining biases. These benefits are primarily realized due to the existence of a consistent and unbiased estimator of the Sharpe ratio.

6 Robustness Tests

In this section, we investigate whether our key findings continue to hold when we consider: (i) different VIX thresholds to classify normal and high uncertainty regimes over the out-of-sample period; (ii) recursive windows of VIX observations (i.e., using the entire available history of VIX observations) to classify economic uncertainty regimes over the out-of-sample period. Our primary focus is to check whether *OPT_COMB* – the 2-tuple predictor combination [*EP*, *INFL*] that delivers the highest forecast accuracy in the baseline analysis – continues to outperform the 14 individual predictors, the historical average ERP, and the mean combination forecast in these tests. If this is indeed the case, then our findings are robust.

In our baseline analysis, we find that the highest forecast accuracy over the holdout sample period is generated by choosing *EP* during periods of normal economic uncertainty and *INFL* during periods of high economic uncertainty. Accordingly, *OPT_COMB* is the 2-tuple predictor combination [*EP*, *INFL*]. By construction, a different combination of predictors may deliver the highest forecast accuracy over the holdout sample period when we choose different VIX thresholds or use recursive windows of VIX observations to classify economic uncertainty regimes.[24] However, in these robust-

[24] We consider three different VIX thresholds to classify economic uncertainty regimes

ness tests, we are interested in whether *OPT_COMB* continues to deliver superior forecast accuracy relative to the individual and benchmark predictors. If this is the case, then we can propose a definitive forecasting strategy that is useful to investors in real time. More specifically, investors will be able to generate superior forecasts if they first classify the one-month-ahead level of economic uncertainty into one of the two regimes and then select forecasts based on *EP* during periods of normal economic uncertainty and *INFL* during periods of high economic uncertainty.

6.1 Different VIX Thresholds

In the analysis that has been discussed so far, we employ the baseline VIX threshold of 75^{th} percentile, where the 75^{th} percentile of in-sample VIX observations is used to classify each month in the out-of-sample period as belonging to the normal or high economic uncertainty regime. In this robustness test, we set January 2005 to December 2020 as the forecast evaluation period (the same as our earlier analysis) and consider two relatively more extreme VIX thresholds to classify economic uncertainty regimes over the out-of-sample period: (i) a threshold of 80^{th} percentile, where the 80^{th} percentile of in-sample VIX observations is used to classify each month in the out-of-sample period as belonging to the normal or high economic uncertainty regime; (ii) a more extreme threshold of 85^{th} percentile, where the 85^{th} percentile of in-sample VIX observations is used to classify each month in the out-of-sample period as belonging to the normal or high economic uncertainty regime.[25] Our motivation for considering multiple VIX thresholds

over the out-of-sample period. Moreover, for each of these choices of VIX thresholds, we use either rolling or recursive windows of VIX observations to classify the economic uncertainty regimes. Therefore, there are a total of six possible regime-classification approaches, one of which corresponds to our baseline analysis. Interestingly, the 2-tuple predictor combination [*EP*, *INFL*] delivers the highest forecast accuracy over the holdout sample period in four out of the six scenarios. In the remaining two scenarios, the forecast accuracy of the predictor combination is close to the best, outperforming both the individual and benchmark predictors.

[25]Section 2 provides a detailed discussion on the regime classification procedure.

is to establish that the forecast accuracy of *OPT_COMB* is not overly reliant on subjective thresholds.

Table 8 reports the R^2_{OS} statistic, the utility gain, the Sharpe ratio, and the haircut Sharpe ratio of select predictors for the different choices of VIX thresholds (including the baseline VIX threshold). Panel A reports the statistics for the benchmark predictors, *HIST_MEAN* and *MEAN_COMB*, while Panel B reports the statistics for *OPT_COMB*.[26] To estimate utility gains and Sharpe ratios, we assume that the investor has a power utility function with CRRA equal to 2.[27] We find that *OPT_COMB* outperforms the historical average ERP over all the three VIX threshold choices, as evidenced by its positive and statistically significant R^2_{OS} statistics (third column of Panel B). *OPT_COMB* consistently outperforms the mean combination forecast also, as the R^2_{OS} statistic of the latter predictor is negative and statistically insignificant (third column of Panel A). The superior forecast accuracy is economically significant since the utility gains (fourth column of Panel B) and the Sharpe ratios (fifth and sixth columns of Panel B) generated by *OPT_COMB* are consistently higher than those generated by the mean combination forecast and the historical average ERP. Overall, the results show that the outperformance of *OPT_COMB* does not depend on the choice of VIX thresholds.

6.2 Recursive Windows of VIX observations

In this robustness test, we use recursive windows of VIX observations – instead of the rolling windows used in our baseline analysis – to classify economic uncertainty regimes over the out-of-sample period. To do this, we

[26]The performance statistics for the 14 individual predictors, the historical average ERP, and the mean combination forecast do not vary with the choice of VIX thresholds and are the same as reported in Table 5. For brevity, we only report the statistics for the historical average ERP and the mean combination forecast. Moreover, there are no entries corresponding to the R^2_{OS} statistic and the utility gain for the historical average ERP. This is because these statistics are designed to measure forecast performance relative to the historical average ERP.

[27]The results are qualitatively similar for CRRA = 1 and CRRA = 3.

follow the same approach for regime-classification as outlined in Section 2, except that we use the entire available history of in-sample VIX observations as the basis for predicting the one-month-ahead economic uncertainty regime. However, if investors form expectations of future economic uncertainty based on recent experience, the recursive window approach may fail to reflect their expectations accurately.

We set January 2005 to December 2020 as the forecast evaluation period and 75^{th} percentile as the VIX threshold, where the 75^{th} percentile of in-sample VIX observations is used to classify each month in the out-of-sample period as belonging to the normal or high economic uncertainty regime. Table 9 reports the R^2_{OS} statistic, the utility gain, the Sharpe ratio, and the haircut Sharpe ratio of select predictors under the rolling and the recursive window approaches for regime-classification. Panel A reports the statistics for the benchmark predictors, *HIST_MEAN* and *MEAN_COMB*, while Panels B and C report the statistics for *OPT_COMB* under the rolling and recursive window approaches, respectively.[28] To estimate utility gains and Sharpe ratios, we assume that the investor has a power utility function with CRRA equal to 2.[29]. We find that *OPT_COMB* outperforms the historical average ERP by a similar margin under both approaches, as evidenced by the nearly equal, positive, and statistically significant R^2_{OS} statistics (third column of Panels B and C). Furthermore, *OPT_COMB* generates marginally higher risk-adjusted returns under the rolling window approach. We can see this by comparing the utility gains, the Sharpe ratios, and the haircut Sharpe ratios in the last three columns of Panels B and C. These findings imply that, regardless of whether investors use the most recent history or the complete history of in-sample VIX data to predict economic uncertainty regimes, they perform comparably when employing the forecasting strategy based on *OPT_COMB*.[30]

[28]For the 14 individual predictors, the historical average ERP and the mean combination forecast, the results are identical under both the regime-classification approaches.

[29]The results are qualitatively similar for CRRA = 1 and CRRA = 3.

[30]In unreported results, we find that *OPT_COMB* also outperforms the 14 individual

7 Conclusion

The asset pricing literature has identified several economic predictors that can predict the in-sample equity risk premium with reasonable success. However, these individual predictors fail to deliver superior out-of-sample forecasts relative to the historical average equity risk premium and the mean combination forecast. The lack of out-of-sample predictive ability can be attributed to structural changes in the return-generating process of the equity risk premium.

We argue that the level of economic uncertainty is one of the key parameters that drives structural changes in the return-generating process of the equity risk premium; the forecast accuracy of any particular predictor depends on the level of economic uncertainty. Consequently, we propose a novel predictor – the optimal combination of predictors – that generates forecasts based on the earnings-price ratio during periods of normal economic uncertainty and inflation during periods of high economic uncertainty. We find that the optimal combination of predictors delivers the highest forecast accuracy relative to the historical average over the out-of-sample period. Neither the mean combination forecast nor the individual predictors outperforms the historical average over the same period.

Our findings underscore the importance of accounting for state dependence in equity risk premium prediction. The VIX acts as a regime-defining variable, capturing shifts in economic uncertainty that alter the performance of ERP models. Similarly, during periods of elevated economic uncertainty, inflation becomes a more informative signal of risk compensation than the earnings-price ratio, which tends to perform better in more stable conditions. This conditional predictability aligns with consumption-based asset pricing models, where the marginal value of information varies across states of the world. Investors can improve return forecasts by conditioning on economic uncertainty, allowing for a more accurate assessment of time-varying

predictors in the robustness tests. These results are available upon request.

risk premia. From a policy perspective, our results emphasize the importance of tailoring monetary policy to the level of economic uncertainty. In low-uncertainty periods, policy influences the equity risk premium through its effects on fundamentals. In contrast, during high-uncertainty periods, inflation drives return dynamics, and overly accommodative policy can increase inflation risk premium and market volatility, potentially undermining stabilization efforts.

We demonstrate the economic significance of superior forecast accuracy by estimating the utility gains and the Sharpe ratios generated by the predictors when its forecasts are utilized through a simple trading strategy. We find that the optimal combination of predictors generates the highest utility gain and Sharpe ratio. Finally, we show that the outperformance of the optimal combination of predictors is robust, irrespective of the approach used to classify the level of economic uncertainty.

Although complex machine learning methods – such as regression trees and neural networks – can match the out-of-sample performance of our optimal predictor combination, they fall short in offering economic interpretability on the relationship between the equity risk premium and the predictor variables. In contrast, the algorithm for identifying the optimal combination of predictors is economically motivated, tractable, and generalizable to other developed and emerging markets, although the optimal predictor combination may change. Moreover, our economic-intuition based model is useful for industry stakeholders who require estimates of the equity risk premium at a higher frequency, such as portfolio managers who rebalance their portfolios frequently and corporate managers who assess short-term investments.

The extant literature on the predictability of equity risk premium focuses on the empirical performance of predictors, rather than examining the potential explanations of predictive ability. Consequently, future research could provide a more complete understanding of the sources of superior predictive ability of our economic uncertainty-conditioned predictor combination.

References

Baillie, R.T., DeGennaro, R.P., 1990. Stock returns and volatility. Journal of Financial and Quantitative Analysis , 203–214.

Ball, R., 1978. Anomalies in relationships between securities' yields and yield-surrogates. Journal of Financial Economics 6, 103–126.

Bansal, N., Stivers, C., 2025. Predicting the equity premium with a high-threshold risk level and the price of risk. Financial Management 54, 123–145.

Bansal, R., Yaron, A., 2004. Risks for the long run: A potential resolution of asset pricing puzzles. The Journal of Finance 59, 1481–1509.

Bao, Y., 2009. Estimation risk-adjusted sharpe ratio and fund performance ranking under a general return distribution. Journal of Financial Econometrics 7, 152–173.

Bekaert, G., Engstrom, E., 2010. Inflation and the stock market: Understanding the "fed model". Journal of Monetary Economics 57, 278–294.

Bekaert, G., Hoerova, M., 2014. The vix, the variance premium and stock market volatility. Journal of Econometrics 183, 181–192.

Bekaert, G., Hoerova, M., Duca, M.L., 2013. Risk, uncertainty and monetary policy. Journal of Monetary Economics 60, 771–788.

Bernanke, B.S., Kuttner, K.N., 2005. What explains the stock market's reaction to federal reserve policy? The Journal of Finance 60, 1221–1257.

Bollerslev, T., Tauchen, G., Zhou, H., 2009. Expected stock returns and variance risk premia. The Review of Financial Studies 22, 4463–4492.

Boudoukh, J., Richardson, M., Whitelaw, R.F., 2008. The myth of long-horizon predictability. The Review of Financial Studies 21, 1577–1605.

Brandt, M.W., Wang, K.Q., 2003. Time-varying risk aversion and unexpected inflation. Journal of Monetary Economics 50, 1457–1498.

Campbell, J.Y., 1987. Stock returns and the term structure. Journal of Financial Economics 18, 373–399.

Campbell, J.Y., Cochrane, J.H., 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. Journal of Political Economy 107, 205–251.

Campbell, J.Y., Giglio, S., Polk, C., Turley, R., 2018. An intertemporal capm with stochastic volatility. Journal of Financial Economics 128, 207–233.

Campbell, J.Y., Hentschel, L., 1992. No news is good news: An asymmetric model of changing volatility in stock returns. Journal of Financial Economics 31, 281–318.

Campbell, J.Y., Shiller, R.J., 1988a. The dividend-price ratio and expectations of future dividends and discount factors. The Review of Financial Studies 1, 195–228.

Campbell, J.Y., Shiller, R.J., 1988b. Stock prices, earnings, and expected dividends. The Journal of Finance 43, 661–676.

Campbell, J.Y., Shiller, R.J., 2001. Valuation ratios and the long-run stock market outlook: An update. NBER Working Paper 8221.

Campbell, J.Y., Thompson, S.B., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? The Review of Financial Studies 21, 1509–1531.

Campbell, J.Y., Vuolteenaho, T., 2004. Inflation illusion and stock prices. The American Economic Review 94, 19–23.

Castelnuovo, E., Pellegrino, G., Særkjær, L.L., 2023. Uncertainty and the business cycle when inflation is high. Unpublished mimeo .

Clark, T.E., West, K.D., 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138, 291–311.

Cochrane, J., 2007. Financial markets and the real economy. Handbook of the Equity Premium .

Cochrane, J.H., 1999. New facts in finance. Economics 23, 36–58.

Cochrane, J.H., 2008. The dog that did not bark: A defense of return predictability. The Review of Financial Studies 21, 1533–1575.

Cukierman, A., Gerlach, S., 2003. The inflation bias revisited: theory and some international evidence. The Manchester School 71, 541–565.

Dangl, T., Halling, M., 2012. Predictive regressions with time-varying coefficients. Journal of Financial Economics 106, 157–181.

Denk, S., Löffler, G., 2024. Predicting the equity premium with combination forecasts: A reappraisal. The Review of Asset Pricing Studies 14, 545–577.

Dennis, P., Mayhew, S., Stivers, C., 2006. Stock returns, implied volatility innovations, and the asymmetric volatility phenomenon. Journal of Financial and Quantitative Analysis 41, 381–406.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 20, 134–144.

Diebold, F.X., Mariano, R.S., 2002. Comparing predictive accuracy. Journal of Business & Economic Statistics 20, 134–144.

Fama, E.F., 1981. Stock returns, real activity, inflation, and money. The American Economic Review 71, 545–565.

Fama, E.F., French, K.R., 1988. Dividend yields and expected stock returns. Journal of Financial Economics 22, 3–25.

Fama, E.F., French, K.R., 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 25, 23–49.

Fama, E.F., Schwert, G.W., 1977. Asset returns and inflation. Journal of Financial Economics 5, 115–146.

Forbes, K.J., Warnock, F.E., 2012. Capital flow waves: Surges, stops, flight, and retrenchment. Journal of International Economics 88, 235–251.

French, K.R., Schwert, G.W., Stambaugh, R.F., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3–29.

Goyal, A., Welch, I., 2003. Predicting the equity premium with dividend ratios. Management Science 49, 639–654.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. The Review of Financial Studies 33, 2223–2273.

Harvey, C.R., Liu, Y., 2015. Backtesting. The Journal of Portfolio Management 42, 13–28.

Hecht, P., Vuolteenaho, T., 2006. Explaining returns with cash-flow proxies. The Review of Financial Studies 19, 159–194.

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

Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics , 65–70.

Jurado, K., Ludvigson, S.C., Ng, S., 2015. Measuring uncertainty. The American Economic Review 105, 1177–1216.

Kothari, S.P., Shanken, J., 1997. Book-to-market, dividend yield, and expected market returns: A time-series analysis. Journal of Financial Economics 44, 169–203.

Lamont, O., 1998. Earnings and expected returns. The Journal of Finance 53, 1563–1587.

Lintner, J., 1975. Inflation and security returns. The Journal of Finance 30, 259–280.

Miller, R.E., Gehr, A.K., 1978. Sample size bias and sharpe's performance measure: A note. Journal of Financial and Quantitative Analysis 13, 943–946.

Nelson, C.R., 1976. Inflation and rates of return on common stocks. The Journal of Finance 31, 471–483.

Pastor, L., Veronesi, P., 2012. Uncertainty about government policy and stock prices. The Journal of Finance 67, 1219–1264.

Pettenuzzo, D., Timmermann, A., Valkanov, R., 2014. Forecasting stock returns under economic constraints. Journal of Financial Economics 114, 517–553.

Piazzesi, M., Schneider, M., 2007. Equilibrium yield curves. NBER Macroeconomics Annual 21, 389–472.

Pontiff, J., Schall, L.D., 1998. Book-to-market ratios as predictors of market returns. Journal of Financial Economics 49, 141–160.

Rapach, D., Zhou, G., 2013. Forecasting stock returns, in: Handbook of Economic Forecasting. Elsevier. volume 2, pp. 328–383.

Rapach, D.E., Strauss, J.K., Zhou, G., 2010. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. The Review of Financial Studies 23, 821–862.

Rayburn, J., 1986. The association of operating cash flow and accruals with security returns. Journal of Accounting Research , 112–133.

Rozeff, M.S., 1984. Dividend yields are equity risk premiums. Journal of Portfolio Management , 68–75.

Spiegel, M., 2008. Forecasting the equity premium: Where we stand today. The Review of Financial Studies 21, 1453–1454.

Timmermann, A., 2008. Elusive return predictability. International Journal of Forecasting 24, 1–18.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. The Review of Financial Studies 21, 1455–1508.

Westerlund, J., Narayan, P., 2014. A random coefficient approach to the predictability of stock returns in panels. Journal of Financial Econometrics 13, 605–664.

Wolff, D., Neugebauer, U., 2019. Tree-based machine learning approaches for equity market predictions. Journal of Asset Management 20, 273–288.

Tables and Figures

Table 1
Variable Definitions
This table defines the variables that are employed in the predictive model for the equity risk premium.

Variable	Definition
Predicted Variable	
Equity Risk Premium (ERP)	The continuously compounded monthly returns on the S&P 500 index (inclusive of dividends) minus the risk-free rate (3-month Treasury bill rate).
Stock Characteristic-related Predictor Variables	
Dividend-price ratio (DP)	The difference between the log of dividends paid on the S&P 500 index and the log of current-month stock price (the S&P 500 index), where dividends are measured using 12-month moving sums.
Dividend-yield (DY)	The difference between the log of dividends paid on the S&P 500 index and the log of lagged (12-month) stock price, where dividends are measured using 12-month moving sums.
Earnings-price ratio (EP)	The difference between the log of earnings on the S&P 500 index and the log of current-month stock price, where earnings are measured using 12-month moving sums.
Dividend-payout ratio (DE)	The difference between the log of dividends and the log of earnings, where dividends and earnings are measured using 12-month moving sums.
Stock variance ($SVAR$)	The sum of squared daily returns on the S&P 500 index.
Book-to-market ratio (BM)	The ratio of book value to market value for the Dow Jones Industrial Average. For the months from March to December, the ratio is computed by dividing the book value at the end of the previous year by the price at the end of the current month. For the months of January and February, the ratio is computed by dividing the book value at the end of two years ago by the price at the end of the current month (Pontiff and Schall, 1998).

(Continued)

Table 1 – *Continued*

Variable	Definition
Net equity expansion (*NTIS*)	The ratio of 12-month moving sums of net issues by NYSE-listed stocks to total end-of-year market capitalization of NYSE stocks.
Interest Rate-related Predictor Variables	
Treasury bill rate (*TBL*)	The interest rate on the 3-month Treasury bill (secondary market rate).
Long-term yield (*LTY*)	The yield on long-term government bonds.
Long-term returns (*LTR*)	The nominal return on long-term government bonds.
Term spread (*TMS*)	The difference between the long-term yield on government bonds and the Treasury bill rate.
Default yield spread (*DFY*)	The difference between BAA- and AAA-rated corporate bond yields.
Default return spread (*DFR*)	The difference between long-term corporate bond and long-term government bond returns.
Macroeconomic Environment-related Predictor Variables	
Inflation (*INFL*)	This is computed from the Consumer Price Index (All Urban Consumers). Following Welch and Goyal (2008), we lag inflation by two months in our monthly regressions because information on the current month's inflation rate is publicly available in the following month.

Table 2
Summary Statistics
This table reports the summary statistics of the variables used in our analysis. The data is at a monthly frequency from January 1990 to December 2020. The average monthly ERP was 0.070% (i.e., 8.39% per annum) over the entire sample period.

Variables	Min.	25 pct.	Median	Mean	75 pct.	Max.
ERP	-0.1678	-0.0168	0.0106	0.0070	0.0331	0.1289
DP	0.0108	0.0174	0.0194	0.0206	0.0223	0.0392
DY	0.0104	0.0192	0.0211	0.0222	0.0245	0.0403
EP	0.0079	0.0391	0.0459	0.0463	0.0552	0.0769
DE	0.2882	0.3530	0.4120	0.5086	0.5320	3.9730
$SVAR$	0.0002	0.0007	0.0014	0.0028	0.0027	0.0732
BM	0.1205	0.2385	0.2861	0.2861	0.3335	0.5225
$NTIS$	-0.0560	-0.0111	0.0070	0.0042	0.0168	0.0457
TBL	0.0001	0.0023	0.0222	0.0263	0.0485	0.0790
LTY	0.0062	0.0302	0.0485	0.0487	0.0639	0.0924
LTR	-0.1124	-0.0116	0.0077	0.0068	0.0243	0.1443
TMS	-0.0041	0.0113	0.0219	0.0224	0.0339	0.0455
DFY	0.0055	0.0071	0.0089	0.0096	0.0106	0.0338
DFR	-0.0976	-0.0063	0.0005	0.0001	0.0063	0.0737
$INFL$	-0.0192	0.0002	0.0019	0.0020	0.0039	0.0122

Table 3
Out-of-Sample Forecast Accuracy (May 1996 - December 2020)
This table reports the forecast accuracy of the predictors over the out-of-sample period from May 1996 to December 2020. Panels A and B report the results for the 14 individual predictors and *MEAN_COMB*, respectively. *MEAN_COMB* is the equally weighted average of forecasts of the 14 individual predictors. We measure forecast accuracy by the Campbell and Thompson (2008) R^2_{OS} statistic. The null hypothesis is that the expected forecast error of the predictor and the historical average ERP forecast (*HIST_MEAN*) are equal, while the alternative hypothesis is that the expected forecast error of the predictor is lower. The p-values reported in the third and fourth columns are computed from the standard normal and the nonparametric bootstrapped sampling distributions of the MSPE-adjusted statistic, respectively. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A

Predictor	R^2_{OS} (%)	p-Value	
		Standard Normal	Bootstrap
DP	-0.1693	0.3991	0.4080
DY	-0.2046	0.3060	0.3053
EP	-1.8859	0.3660	0.3712
DE	-1.9266	0.6404	0.6590
SVAR	-6.1242	0.6077	0.6231
BM	-0.1900	0.5242	0.5282
NTIS	-0.6831	0.4541	0.4435
TBL	-0.7919	0.6398	0.6403
LTY	-0.9008	0.6971	0.6999
LTR	-0.7224	0.8924	0.8933
TMS	-0.7055	0.8963	0.8951
DFY	-2.0760	0.3468	0.3335
DFR	-3.1958	0.9577	0.9447
INFL	-0.6763	0.5836	0.5733

Panel B

Predictor	R^2_{OS} (%)	p-Value	
		Standard Normal	Bootstrap
MEAN_COMB	-0.5209	0.5957	0.5961

Table 4
Holdout Sample Forecast Accuracy (May 1996 - December 2004)
This table reports the forecast accuracy of the predictors in the two economic uncertainty regimes (normal and high) over the holdout sample period from May 1996 to December 2004. Panels A and B report the results for the 14 individual predictors and *MEAN_COMB*, respectively. *MEAN_COMB* is the equally weighted average of forecasts of the 14 individual predictors. We measure forecast accuracy by the Campbell and Thompson (2008) R^2_{OS} statistic. The null hypothesis is that the expected forecast error of the predictor and the historical average ERP forecast (*HIST_MEAN*) are equal, while the alternative hypothesis is that the expected forecast error of the predictor is lower. Statistical significance is estimated from the nonparametric bootstrapped sampling distributions of the MSPE-adjusted statistic. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A

Predictor	R^2_{OS} (%)	
	Normal Uncertainty	High Uncertainty
DP	2.2302	-2.6723
DY	5.5963*	-4.8904
EP	14.8669***	-3.5953
DE	-3.7067	0.0583
SVAR	-5.4295	-0.3251
BM	0.2360	-1.3646
NTIS	-0.6268	-0.7979
TBL	-1.1323	-1.0171
LTY	-0.5464	-2.1610
LTR	-0.6525	-1.2255
TMS	-2.9891	-0.3791
DFY	-0.8759	-1.2070
DFR	-1.4609	-2.1201
INFL	-2.6005	0.1568

Panel B

Predictor	R^2_{OS} (%)	
	Normal Uncertainty	High Uncertainty
MEAN_COMB	0.7292	-1.0064

Table 5
Out-of-Sample Forecast Accuracy (January 2005 - December 2020)
This table reports the forecast accuracy of the predictors over the out-of-sample period from January 2005 to December 2020. Panels A, B, and C report the results for the 14 individual predictors, *MEAN_COMB*, and *OPT_COMB*, respectively. *MEAN_COMB* is the equally weighted average of forecasts of the 14 individual predictors. *OPT_COMB* generates forecasts based on the earnings-price ratio (*EP*) during periods of normal economic uncertainty and inflation (*INFL*) during periods of high economic uncertainty. We measure forecast accuracy by the Campbell and Thompson (2008) R^2_{OS} statistic. The null hypothesis is that the expected forecast error of the predictor and the historical average ERP forecast (*HIST_MEAN*) are equal, while the alternative hypothesis is that the expected forecast error of the predictor is lower. The p-values reported in the third and fourth columns are computed from the standard normal and the nonparametric bootstrapped sampling distributions of the MSPE-adjusted statistic, respectively. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A

Predictor	R^2_{OS} (%)	p-Value	
		Standard Normal	Bootstrap
DP	0.3709	0.2584	0.2624
DY	0.4935	0.1082	0.1071
EP	-5.1318	0.6625	0.6731
DE	-2.3843	0.6188	0.6277
SVAR	-8.8713	0.5860	0.6069
BM	0.2309	0.2534	0.2539
NTIS	-0.6454	0.3851	0.3800
TBL	-0.6102	0.5689	0.5720
LTY	-0.4246	0.5105	0.5126
LTR	-0.5152	0.6720	0.6780
TMS	-0.3047	0.8112	0.8120
DFY	-2.7498	0.3152	0.3087
DFR	-4.0895	0.9015	0.8898
INFL	-0.5869	0.4763	0.4493

Panel B

Predictor	R^2_{OS} (%)	p-Value	
		Standard Normal	Bootstrap
MEAN_COMB	-0.6037	0.5612	0.5670

Panel C

Predictor	R^2_{OS} (%)	p-Value	
		Standard Normal	Bootstrap
OPT_COMB	1.3272*	0.0579	0.0681

Table 6
Economic Significance of Forecast Accuracy – Utility gains
This table reports the annualized utility gain (relative to historical average ERP) realized by a mean-variance investor who rebalances their portfolio monthly between the market portfolio (the S&P 500) and risk-free 30-day T-bills based on out-of-sample forecasts of the predictors. We restrict the weight on the market portfolio to lie between 0 and 1.5 each month, which prevents the investor from shorting the market portfolio or leveraging more than 50%. For computing portfolio weights, we assume that the investor has a power utility function with the coefficient of relative risk aversion (CRRA) values of 1, 2, and 3. We measure the utility gain from a predictor as the difference between the realized utility from its forecasts and the realized utility from the historical average ERP forecast. The out-of-sample period over which we estimate the utility gains is from January 2005 to December 2020. Panels A, B, and C report the results for the 14 individual predictors, *MEAN_COMB*, and *OPT_COMB* respectively. *MEAN_COMB* is the equally weighted average of forecasts of the 14 individual predictors. *OPT_COMB* generates forecasts based on the earnings-price ratio (*EP*) during periods of normal economic uncertainty and inflation (*INFL*) during periods of high economic uncertainty.

Panel A

Predictor	Utility Gain w.r.t. Historical Average ERP (%)		
	CRRA=1	CRRA=2	CRRA=3
DP	0.0000	1.0986	0.0002
DY	0.0349	1.0575	1.3609
EP	0.0894	3.1092	2.9372
DE	-0.1694	0.7931	1.8759
SVAR	-0.8320	0.3577	-0.2180
BM	0.0000	0.6606	-0.4669
NTIS	-0.6960	0.4933	-0.0693
TBL	-2.2998	-0.1646	1.1560
LTY	-3.1996	-0.4897	0.9157
LTR	-0.4159	-1.1001	-1.4173
TMS	-0.7019	-0.4806	0.1830
DFY	0.9913	1.9820	2.0793
DFR	-2.7741	-0.9625	0.2986
INFL	-1.3274	-1.1283	-1.6800

Panel B

Predictor	Utility Gain w.r.t. Historical Average ERP (%)		
	CRRA=1	CRRA=2	CRRA=3
MEAN_COMB	-1.2562	-0.5715	0.4119

Panel C

Predictor	Utility Gain w.r.t. Historical Average ERP (%)		
	CRRA=1	CRRA=2	CRRA=3
OPT_COMB	1.9022	3.9620	2.7450

Table 7
Economic Significance of Forecast Accuracy – Haircut Sharpe Ratios
This table reports the annualized Sharpe ratios realized by a mean-variance investor who rebalances their portfolio monthly between the market portfolio (the S&P 500) and risk-free 30-day T-bills based on out-of-sample forecasts of the predictors. We restrict the weight on the market portfolio to lie between 0 and 1.5 each month, which prevents the investor from shorting the market portfolio or leveraging more than 50%. For computing portfolio weights, we assume that the investor has a power utility function with the coefficient of relative risk aversion (CRRA) value of 2. The Sharpe ratio associated with a predictor is the ratio of the sample mean and the sample standard deviation of the monthly excess returns (over the risk-free rate) generated from utilizing its forecasts through the trading strategy. To address the issue of data mining, we follow the multiple hypothesis testing framework proposed by Holm (1979) as it is designed to eliminate all false discoveries. The framework provides an adjustment to the p-value of a single test (p^S) to obtain the multiple-testing-adjusted p-value (p^M). We then estimate the haircut Sharpe ratio as the imputed Sharpe ratio corresponding to the multiple testing p-value p^M. The out-of-sample period over which we estimate the Sharpe ratios is from January 2005 to December 2020. Panel A reports the annualized Sharpe ratios ($\widehat{SR}$), p^S, p^M, and the haircut Sharpe ratios ($\widehat{HSR}$) for the 14 individual predictors and the historical average ERP ($HIST\_MEAN$). Panels B and C report the same statistics for $MEAN\_COMB$ and OPT_COMB, respectively. $MEAN\_COMB$ is the equally weighted average of forecasts of the 14 individual predictors. OPT_COMB generates forecasts based on the earnings-price ratio (EP) during periods of normal economic uncertainty and inflation ($INFL$) during periods of high economic uncertainty. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A

Predictor	$\widehat{SR}$ (Annual)	p^S	p^M	$\widehat{HSR}$ (Annual)
DP	0.5983	0.0195	0.2052	0.2299
DY	0.6054	0.0171	0.2052	0.2246
EP	0.7315	0.0040	0.0680	0.4061*
DE	0.6170	0.0145	0.1885	0.2425
SVAR	0.5840	0.0280	0.2052	0.2360
BM	0.5733	0.0233	0.2052	0.2268
NTIS	0.6246	0.0109	0.1526	0.2765
TBL	0.5805	0.0182	0.2052	0.2173
LTY	0.5612	0.0210	0.2052	0.2166
LTR	0.4947	0.0399	0.2052	0.2269
TMS	0.5336	0.0338	0.2052	0.2311
DFY	0.7066	0.0056	0.0840	0.3782*
DFR	0.5131	0.0367	0.2052	0.2254
INFL	0.5030	0.0372	0.2052	0.2287
HIST_MEAN	0.5505	0.0288	0.2052	0.2292

Panel B

Predictor	$\widehat{SR}$ (Annual)	p^S	p^M	$\widehat{HSR}$ (Annual)
MEAN_COMB	0.5396	0.0343	0.2052	0.2295

Panel C

Predictor	$\widehat{SR}$ (Annual)	p^S	p^M	$\widehat{HSR}$ (Annual)
OPT_COMB	0.7552	0.0048	0.0768	0.4110*

Table 8
Robustness Tests – Different VIX Thresholds
This table reports the R^2_{OS} statistic, the utility gain (relative to historical average ERP), the Sharpe ratio ($\widehat{SR}$), and the haircut Sharpe ratio ($\widehat{HSR}$) of select predictors for different choices of VIX thresholds. Panel A reports the statistics for the benchmark predictors, *HIST_MEAN* and *MEAN_COMB*, while Panel B reports the statistics for *OPT_COMB*. *MEAN_COMB* is the equally weighted average of forecasts of the 14 individual predictors. *OPT_COMB* generates forecasts based on the earnings-price ratio (*EP*) during periods of normal economic uncertainty and inflation (*INFL*) during periods of high economic uncertainty. To estimate utility gains and Sharpe ratios, we assume that the investor has a power utility function with the coefficient of relative risk aversion (CRRA) value of 2. We set January 2005 to December 2020 as the forecast evaluation period and consider three different Volatility Index (VIX) thresholds to classify the economic uncertainty regime for each month of the out-of-sample period: (i) the baseline threshold of 75^{th} percentile, where the 75^{th} percentile of in-sample VIX observations is used to classify each month in the out-of-sample period as belonging to the normal or high economic uncertainty regime; (ii) a more extreme threshold of 80^{th} percentile, where the 80^{th} percentile of in-sample VIX observations is used to classify each month in the out-of-sample period as belonging to the normal or high economic uncertainty regime; (iii) an even more extreme threshold of 85^{th} percentile, where the 85^{th} percentile of in-sample VIX observations is used to classify each month in the out-of-sample period as belonging to the normal or high economic uncertainty regime. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A					
Predictor	VIX Threshold	R^2_{OS} (%)	Utility Gain w.r.t. Historical Average ERP (%)	$\widehat{SR}$ (Annual)	$\widehat{HSR}$ (Annual)
HIST MEAN	-	-	-	0.5505**	0.2292
MEAN COMB	-	-0.6037	-0.5715	0.5396**	0.2295
Panel B					
Predictor	VIX Threshold	R^2_{OS} (%)	Utility Gain w.r.t. Historical Average ERP (%)	$\widehat{SR}$ (Annual)	$\widehat{HSR}$ (Annual)
OPT_COMB	75^{th} percentile	1.3272*	3.9620	0.7552***	0.4110*
OPT_COMB	80^{th} percentile	1.1950*	3.7213	0.7416***	0.3903*
OPT_COMB	85^{th} percentile	1.4160*	3.8795	0.7473***	0.4034*

Table 9
Robustness Tests – Recursive Window Approach for regime-classification
This table reports the R^2_{OS} statistics, the utility gains (relative to historical average ERP), the Sharpe ratios ($\widehat{SR}$) and the haircut Sharpe ratios ($\widehat{HSR}$) of select predictors under the rolling and the recursive window approaches for regime-classification. Panel A reports the statistics for the benchmark predictors, *HIST_MEAN* and *MEAN_COMB*, while Panels B and C report the statistics for *OPT_COMB* under the rolling and recursive window approaches, respectively. *MEAN_COMB* is the equally weighted average of forecasts of the 14 individual predictors. *OPT_COMB* generates forecasts based on the earnings-price ratio (*EP*) during periods of normal economic uncertainty and inflation (*INFL*) during periods of high economic uncertainty. To estimate utility gains and Sharpe ratios, we assume that the investor has a power utility function with the coefficient of relative risk aversion (CRRA) value of 2. Under the rolling (recursive) window approach, we use the recent (entire) available history of in-sample Volatility Index (VIX) observations as the basis for predicting the one-month-ahead economic uncertainty regime. We set January 2005 to December 2020 as the forecast evaluation period and 75^{th} percentile as the VIX threshold, where the 75^{th} percentile of in-sample VIX observations is used to classify each month in the out-of-sample period as belonging to the normal or high economic uncertainty regime. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A					
Predictor	VIX Threshold	R^2_{OS} (%)	Utility Gain w.r.t. Historical Average ERP (%)	$\widehat{SR}$ (Annual)	$\widehat{HSR}$ (Annual)
HIST MEAN	-	-	-	0.5505**	0.2292
MEAN COMB	-	-0.6037	-0.5715	0.5396**	0.2295
Panel B: Rolling windows of VIX observations					
Predictor	VIX Threshold	R^2_{OS} (%)	Utility Gain w.r.t. Historical Average ERP (%)	$\widehat{SR}$ (Annual)	$\widehat{HSR}$ (Annual)
OPT_COMB	75^{th} percentile	1.3272*	3.9620	0.7552***	0.4110*
Panel C: Recursive windows of VIX observations					
Predictor	VIX Threshold	R^2_{OS} (%)	Utility Gain w.r.t. Historical Average ERP (%)	$\widehat{SR}$ (Annual)	$\widehat{HSR}$ (Annual)
OPT_COMB	75^{th} percentile	1.3355*	3.1821	0.7208***	0.3762*

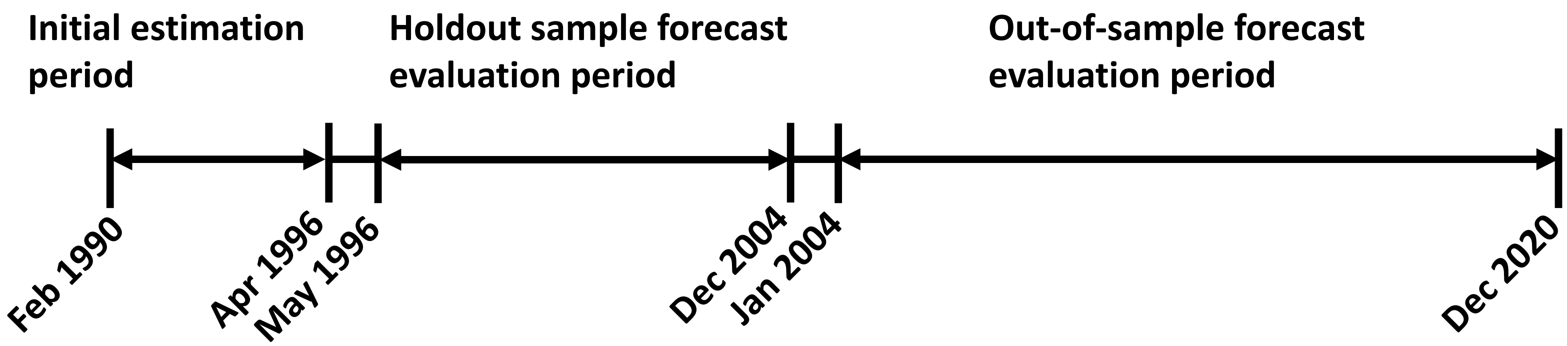
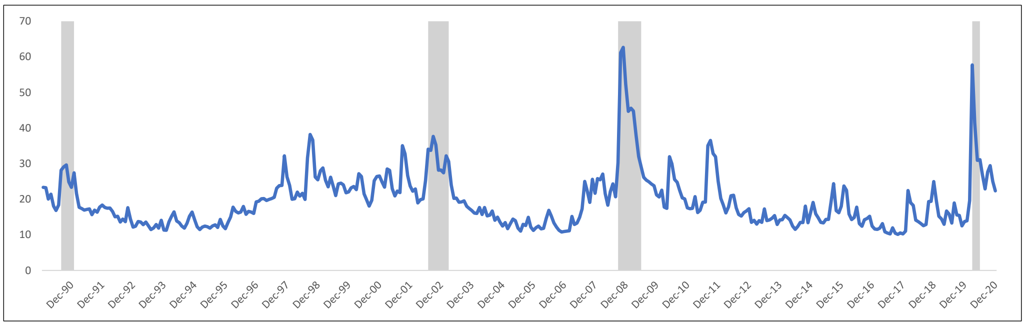
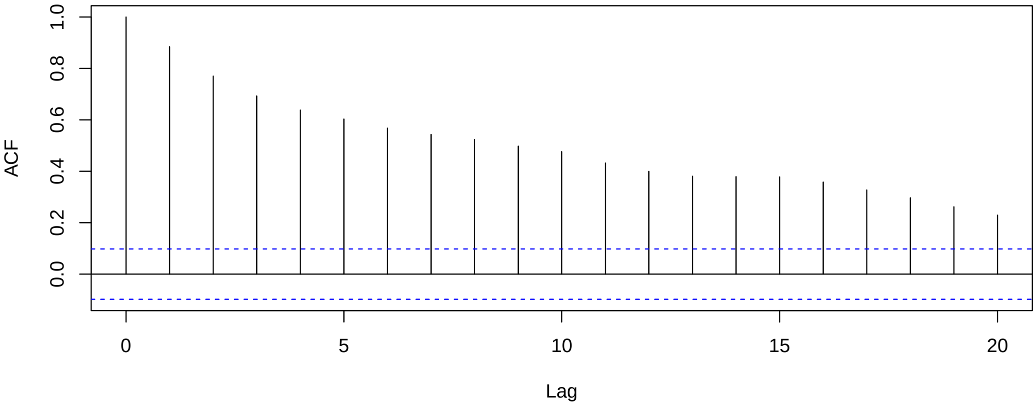


Figure 1
Forecast Timeline
This figure highlights the distinction between the initial estimation period that expands recursively one month at a time, the holdout sample period for identifying the optimal combination of predictors (OPT_COMB), and the out-of-sample period for evaluating the forecast accuracy of predictors. The recursive estimation process generates one-month-ahead ERP forecasts from May 1996 to December 2020. This period is then partitioned into the holdout sample forecast evaluation period and the out-of-sample forecast evaluation period.



Panel A: Volatility Index (VIX)



Panel B: Autocorrelation Function (ACF)

Figure 2
Volatility Index
Panel A of this figure depicts the evolution of economic uncertainty from January 1990 to December 2020. We measure economic uncertainty by the Volatility Index (VIX) – an estimate of constant, 30-day expected volatility of the US equity market that is imputed from the prices of S&P 500 index options. The shaded vertical bands denote economic recessions as estimated by the National Bureau of Economic Research (NBER). Panel B plots the autocorrelation function (ACF) of the VIX observations.

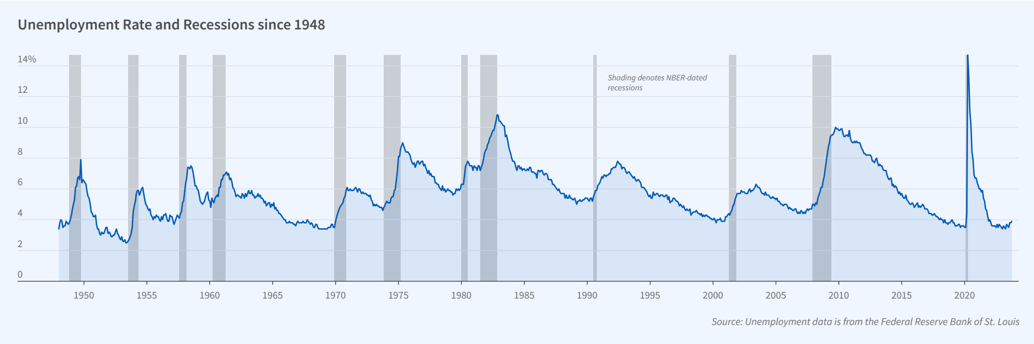
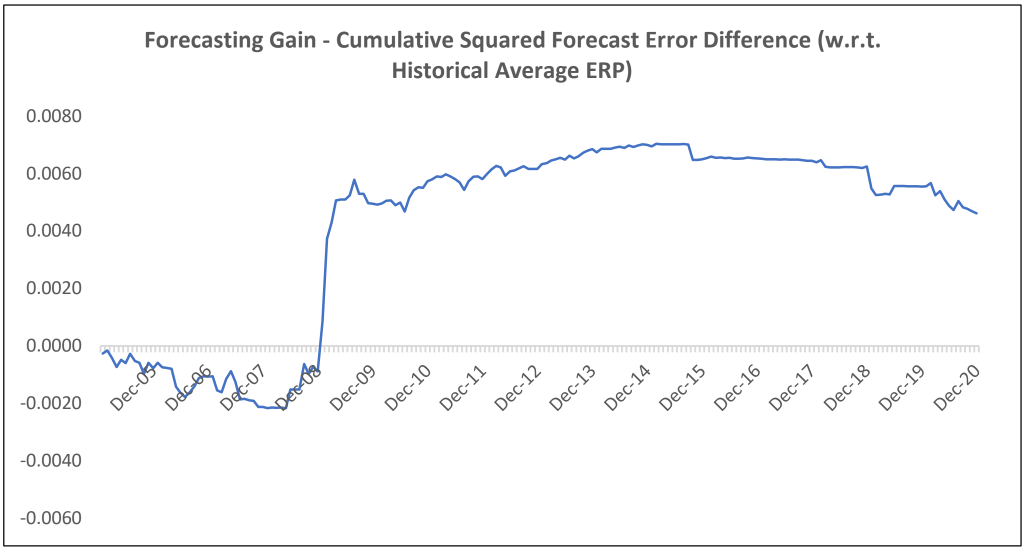
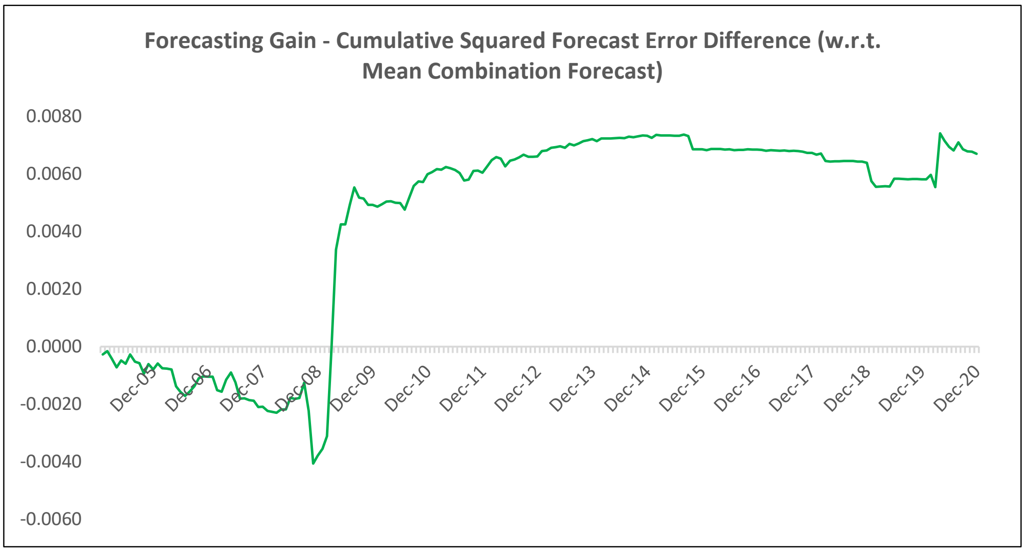


Figure 3
NBER-dated Business Cycles
This figure tracks the economic expansions and recessions that have occurred in the US since 1948. Periods of low (high) unemployment rate indicate expansions (recessions). The shaded vertical bands denote economic recessions as estimated by the National Bureau of Economic Research (NBER).

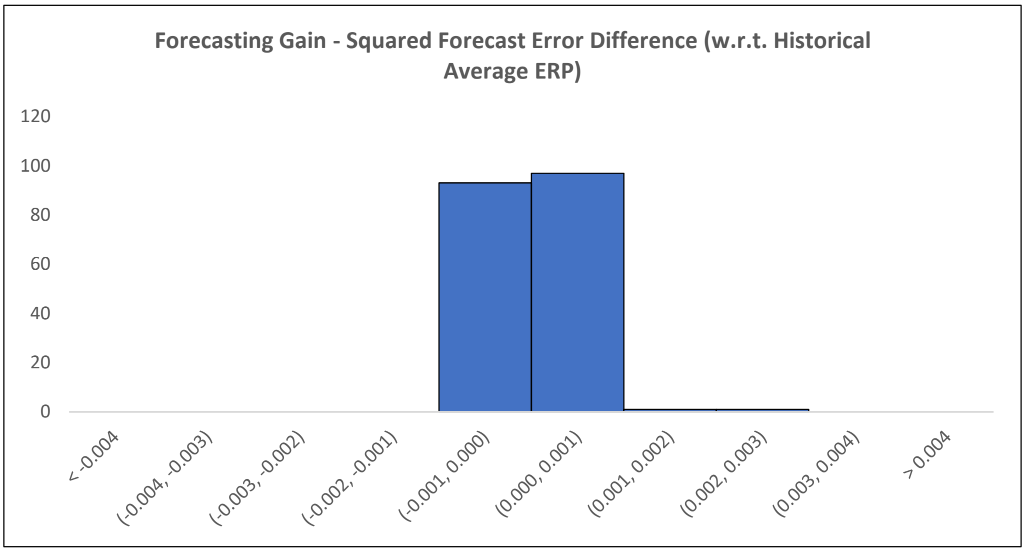


Panel A: HIST_MEAN minus OPT_COMB

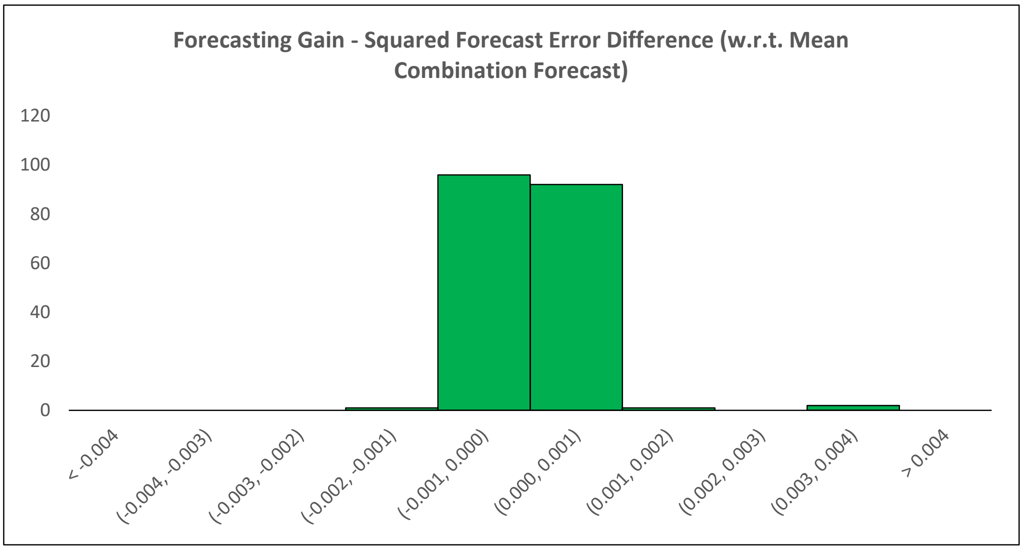


Panel B: MEAN_COMB minus OPT_COMB

Figure 4
Cumulative Squared Forecast Error for the Optimal Combination of Predictors relative to the Historical Average Equity Risk Premium and the Mean Combination Forecast
This figure illustrates the forecasting gains of the optimal combination of predictors (*OPT_COMB*) relative to the benchmark predictors – the historical average equity risk premium (*HIST_MEAN*) and the mean combination forecast (*MEAN_COMB*) – by plotting the time series of the differences between cumulative squared forecast error for the benchmark predictors and the cumulative squared forecast error for *OPT_COMB*. The out-of-sample period is from January 2005 to December 2020. Panel A plots the difference between the cumulative squared forecast error for *HIST_MEAN* and the cumulative squared forecast error for *OPT_COMB*. Panel B plots the difference between the cumulative squared forecast error for *MEAN_COMB* and the cumulative squared forecast error for *OPT_COMB*.



Panel A: HIST_MEAN minus OPT_COMB



Panel B: MEAN_COMB minus OPT_COMB

Figure 5
Histogram of Squared Forecast Error for the Optimal Combination of Predictors relative to the Historical Average Equity Risk Premium and the Mean Combination Forecast
This figure plots the histogram of the difference between squared forecast error for the benchmark predictors – the historical average equity risk premium (*HIST_MEAN*) and the mean combination forecast (*MEAN_COMB*) – and the squared forecast error for the optimal combination of predictors (*OPT_COMB*). The out-of-sample period is from January 2005 to December 2020. Panel A plots the histogram of the difference in squared forecast error for *HIST_MEAN* and the squared forecast error for *OPT_COMB*. The mean (median) of the histogram in Panel A is 2.4×10^{-5} (6.2×10^{-7}). Panel B plots the histogram of the difference in squared forecast error for *MEAN_COMB* and the squared forecast error for *OPT_COMB*. The mean (median) of the histogram in Panel B is 3.5×10^{-5} (6.0×10^{-7}).

Appendix A

Table A.1
Robustness Tests – Alternative Holdout Sample Forecast Accuracy (May 1996 - December 2014)
This table reports the forecast accuracy of the predictors in the two economic uncertainty regimes (normal and high) over the longer holdout sample period from May 1996 to December 2014. Panels A and B report the results for the 14 individual predictors and *MEAN_COMB*, respectively. *MEAN_COMB* is the equally weighted average of forecasts of the 14 individual predictors. We measure forecast accuracy by the Campbell and Thompson (2008) R^2_{OS} statistic. The null hypothesis is that the expected forecast error of the predictor and the historical average ERP forecast (*HIST_MEAN*) are equal, while the alternative hypothesis is that the expected forecast error of the predictor is lower. Statistical significance is estimated from the nonparametric bootstrapped sampling distributions of the MSPE-adjusted statistic. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A

Predictor	R^2_{OS} (%) Normal Uncertainty	R^2_{OS} (%) High Uncertainty
DP	1.5056*	-1.8150
DY	3.4841**	-2.4859
EP	7.2619***	-8.7354
DE	-2.2919	-2.7277
SVAR	-2.0580	-1.8440
BM	0.9467*	-1.2187
NTIS	-0.0834	-0.3831
TBL	-0.9491	-1.2599
LTY	-0.7164	-1.7455
LTR	-0.5418	-1.2546
TMS	-1.6158	-0.3644
DFY	0.2440	-4.2278
DFR	-1.5766	-4.7909
INFL	-4.7564	2.3833**

Panel B

Predictor	R^2_{OS} (%) Normal Uncertainty	R^2_{OS} (%) High Uncertainty
MEAN_COMB	0.3781	-0.7304

Appendix B

The Sharpe ratio ($\hat{\zeta}$) of the trading strategy based on the ERP forecasts of a given predictor is:[31]

$$\hat{\zeta} = \frac{\hat{\mu}}{\hat{\sigma}} \tag{B.1}$$

where $\hat{\mu}$ and $\hat{\sigma}$ are the sample mean and sample standard deviation of the monthly excess returns from the strategy.

Under the assumption that the excess returns are independently and identically distributed (i.i.d) samples from a normal distribution, $\hat{\zeta}$ is biased and the bias is proportional to the population Sharpe ratio (Miller and Gehr, 1978). The bias factor, $\frac{\mathbb{E}[\hat{\zeta}]}{\zeta}$, is more than 1 and converges to 1 as the sample size, n, increases.[32] The bias is negligible for our sample size of 192 excess returns from January 2005 to December 2020.

Relaxing the assumption of normally distributed excess returns, but maintaining the i.i.d assumption, Bao (2009) proposes the following *approximately unbiased* estimator of ζ:

$$\hat{\zeta}_{ab} = \hat{\zeta} - \frac{3}{4n}\hat{\zeta} + \frac{1}{2n}\gamma_1 - \frac{3}{8n}\hat{\zeta}\gamma_2 \tag{B.2}$$

where γ_1 and γ_2 are the Pearson's measures of skewness and kurtosis. For a normal distribution, these parameters are all equal to zero.[33] The estimator is *approximately unbiased* because the exact formulae for the expectation of the estimator (equation (B.2)) contain higher order terms of the sample size, n. We ignore these higher-order terms since our relatively large sample size of 192 observations introduces minimal approximation error in the computations while reducing computational complexity significantly. Since $\hat{\zeta}_{ab}$ depends on unknown population parameters (γ_1 and γ_2), we replace them

[31] We drop subscript i, which indexes the predictor, for notational simplicity.
[32] For example, the bias factor is 1.08 for n=12, 1.02 for n=40, and 1.01 for n=75.
[33] We find that the excess returns generated by the predictors are not normally distributed (using the Jarque-Bera test).

with their corresponding sample estimates. The sample estimates of γ_1 and γ_2 are the Fisher's k-statistics.

When excess returns are not normally distributed, the bias-corrected estimator of the Sharpe ratio ($\hat{\zeta}_{ab}$) has no known theoretical probability distribution. Hence, we generate the sampling distribution of the estimator by bootstrapping the excess return series. However, the estimator is still likely to be biased because of the assumption of i.i.d excess returns, which is unlikely to hold in real world data.[34] If the estimator is *approximately unbiased*, then the mean of the bootstrapped sampling distribution should be very close to the original sample statistic $\bar{\zeta}_{ab}$.[35] Therefore, we estimate the bias ($\hat{B}_{ab}$) as the difference between the mean of the bootstrapped sampling distribution and the original sample statistic $\bar{\zeta}_{ab}$. Post these adjustments, we obtain the following unbiased estimator of the population Sharpe ratio under a general return-generating process of excess returns:

$$\hat{\zeta}_{gdgp} = \hat{\zeta}_{ab} - \hat{B}_{ab} \tag{B.3}$$

To conduct hypothesis tests, we first generate the sampling distribution of $\hat{\zeta}_{gdgp}$ by simply shifting the bootstrapped sampling distribution of $\hat{\zeta}_{ab}$ by the bias $\hat{B}_{ab}$.[36] Then, from the sampling distribution of $\hat{\zeta}_{gdgp}$, we calculate the area in the right tail (one-sided p-value) corresponding to the original sample statistic $\bar{\zeta}_{gdgp}$ ($\bar{\zeta}_{ab} - \hat{B}_{ab}$). This enables us to compute the statistical significance of the estimated Sharpe ratio.

[34] Using the Ljung-Box test, we are unable to reject the joint null hypothesis of no autocorrelation in excess returns (up to 5 lags) for all the predictors. However, the Ljung-Box test examines whether the excess return series is uncorrelated rather than i.i.d, which is a stronger assumption than uncorrelatedness.

[35] This is because bootstrapping treats the sample as the population, which implies that the value of the population parameter is equal to the original sample statistic $\bar{\zeta}_{ab}$.

[36] We shift the bootstrapped sampling distribution of $\hat{\zeta}_{ab}$ to the left (right) if the sign of the bias, $\hat{B}_{ab}$, is positive (negative).