Review of Finance, 2023, 619–657
https://doi.org/10.1093/rof/rfac035
Advance Access Publication Date: 19 May 2022

OXFORD

On the Economic Significance of Stock Return Predictability*

Scott Cederburg[1], Travis L. Johnson[2], and Michael S. O'Doherty[3]

[1]Eller College of Management, University of Arizona, USA, [2]McCombs School of Business, University of Texas at Austin, USA and [3]Trulaske College of Business, University of Missouri, USA

Abstract

We study the effects of time-varying volatility and investment horizon on the economic significance of stock market return predictability from the perspective of Bayesian investors. Using a vector autoregression framework with stochastic volatility (SV) in market returns and predictor variables, we assess a broad set of twenty-six predictors with both in-sample and out-of-sample designs. Volatility and horizon are critically important for assessing return predictors, as these factors affect how an investor learns about predictability and how she chooses to invest based on return forecasts. We find that statistically strong predictors can be economically unimportant if they tend to take extreme values in high volatility periods, have low persistence, or follow distributions with fat tails. Several popular predictors exhibit these properties such that their impressive statistical results do not translate into large economic gains. We also demonstrate that incorporating SV leads to substantial utility gains in real-time forecasting.

Keywords: Stock market return predictability, Time-varying volatility, Economic significance

JEL classification: G10, G11, G12

Received August 3, 2021; accepted March 17, 2022 by Editor.

1. Introduction

Although investors and academics have long studied whether stock market returns are predictable, many historically important predictor candidates have met with challenges on

* We thank Francisco Barillas, David Brown, Paolo Giordani, Amit Goyal (the editor), Chris Lamoureux, two anonymous referees, participants at the 2019 European Finance Association Annual Meeting and the 2019 Melbourne Asset Pricing Meeting, and seminar participants at the University of Arizona, the University of Nebraska–Lincoln, and the University of Texas at Austin for helpful comments and suggestions. The data underlying this article will be shared on reasonable request to the corresponding author.

two fronts. In-sample predictability tests tend to generate only weak evidence in favor of predictability with marginal statistical significance and low in-sample predictive R^2s. Out-of-sample tests reveal difficulties in return forecasting using information available in real time (Goyal and Welch, 2008). Researchers have recently uncovered new predictors that achieve strong in-sample statistical significance relative to their historical peers, and many of the studies that propose these predictors also demonstrate success with real-time forecasts. Several of these predictor variables do, however, display extreme time-series properties and strong relations with stock market volatility that cast doubt on the plausibility and viability of their return forecasts.

We study in-sample and out-of-sample return predictability evidence from a broad set of twenty-six predictor variables, fourteen from Goyal and Welch (2008) and twelve from more recent publications in top finance journals. A key aspect of our study is that we consider the effects of time-varying stock market volatility on predictability evidence. We do so by taking the perspective of Bayesian investors who learn about the dynamics of returns and predictors from a vector autoregression model in which returns and predictors have stochastic volatility (SV). In addition to considering a 1-month investment horizon, we also investigate multi-month horizons. Within this framework, we examine two types of investors: (i) those who learn from the full time series of data and (ii) those who must learn from the data in real time. Our study thus provides an expansive view of the effects of time-varying volatility and horizon on the in-sample and out-of-sample evidence on return predictability.

We find that considering time-varying volatility and horizon has economically important effects for inferences about predictability. In our in-sample analysis, we demonstrate that predictors with strong statistical support in a traditional ordinary least squares (OLS) setting can be economically unimportant, particularly if they tend to make extreme forecasts in periods of high market volatility, have low persistence, or follow distributions with fat tails. At the same time, some other variables with weak statistical evidence display relatively strong economic significance in models with SV. In our out-of-sample setting, accounting for time-varying volatility while making real-time forecasts leads to greater success. A real-time investor with a 1-month horizon who learns from a constant-volatility (CV) model realizes a utility gain for only eleven of the twenty-six predictors and most of these gains are economically small. The analogous investor who learns from a model with SV, in contrast, realizes real-time utility gains from twenty of the twenty-six predictors. All told, six predictors (two from Goyal and Welch, 2008, and four from newer studies) deliver economically meaningful performance in both in-sample and out-of-sample tests at 1- and 3-month horizons.[1]

Our allowance for time-varying volatility and multi-month horizons differs from the approach in much of the previous literature. In a seminal paper, Kandel and Stambaugh (1996) consider the economic significance of predictability in a setting with CV in stock returns and a 1-month investment horizon, and they show that even weak statistical evidence of predictability can translate into sizeable economic effects. In particular, the current value of a weak predictor can have a strong influence on a Bayesian investor's optimal allocation to stocks and the investor perceives substantial utility gains from investing on predictability evidence. As an example, the OLS slope coefficient for the dividend-price ratio in a monthly forecasting regression has a P-value of 0.16 and the regression R^2 value

1 These six predictors are the Treasury bill yield, the long-term Treasury bond yield, Chen and Zhang's (2011) EG, Cooper and Priestly's (2009) output gap, Jones and Tuzel's (2013) new orders-to-shipments of durable goods, and Huang and Kilic's (2019) gold-to-platinum price ratio.

is only 0.002. When the 1-month, CV investor studied by Kandel and Stambaugh (1996) forms beliefs based on information from the dividend-price ratio, she varies her allocation to stocks between 3% and 100% during the 1927–2017 period and perceives a non-trivial utility gain of 0.24% per year in certainty equivalent return (CER).

We initially show that in-sample CERs in Kandel and Stambaugh's (1996) 1-month, CV setting are tightly linked to the OLS R^2 from a predictive regression,

$$r_{t+1} = \alpha + \beta x_t + \epsilon^r_{t+1}, \tag{1}$$

where r_{t+1} is the log excess stock market return and x_t is the value of a predictor variable at the end of month t. In this sense, the OLS R^2 metric that is commonly emphasized in the predictability literature is economically meaningful. Within this framework, we demonstrate that many of the twenty-six predictors are economically significant from the Bayesian investors' perspective. Among the fourteen Goyal–Welch variables, which tend to have relatively low statistical support, the CER gains range from 0.03% per year for the dividend-earnings ratio to 0.68% per year for net equity expansion. The twelve new predictors tend to produce larger economic benefits, in line with their stronger in-sample statistical significance. The range of annual CER gains is 0.55% for the nearness-to-Dow-historical-high variable of Li and Yu (2012) to 4.36% for the variance risk premium (VRP) of Bollerslev, Tauchen, and Zhou (2009). Panel A of Figure 1 shows that OLS R^2 provides a good fit to CER gain across predictors (the Spearman rank correlation is 0.999), such that the statistical performance of predictors is quite informative about their relative in-sample economic value in the 1-month, CV setting.

We proceed to study return predictability in the context of Bayesian investors who account for time variation in stock market volatility. From a statistical perspective, the evidence in favor of time-varying volatility in stock returns is overwhelming (e.g., Bollerslev, Engle, and Nelson, 1994). From a portfolio choice perspective, several studies document the economic importance of considering time-varying volatility (e.g., Fleming, Kirby, and Ostdiek, 2001, 2003).[2] This evidence suggests that the SV setting provides a sensible backdrop for studying the economic significance of return predictability.

SV interacts with return predictability to produce three primary effects on portfolio choice with a 1-month horizon. First, a statistical effect changes inferences about the statistical strength of a predictor. An investor who believes in SV weights observations in the market return predictability regression based on their precision, such that the investor effectively downweights (upweights) information from high-volatility (low-volatility) periods and more efficiently learns about the predictability relation. Inferences about the strength of a predictor can thus differ across the CV and SV settings.[3] Second, an average volatility effect arises because of the positive time-series skewness in conditional market volatility. In most months, the estimated conditional variance is less than the estimated unconditional variance. The SV investor therefore believes that she can usually trade more aggressively than the CV investor, such that the average volatility effect increases the economic significance of each predictor. Third, a portfolio effect captures interactions between market

2 We also find that our Bayesian investors perceive large utility gains from acknowledging time variation in volatility.

3 In a frequentist setting, Johnson (2019) shows that in-sample and out-of-sample inferences about predictability can change when estimating the predictability regression with weighted least squares (WLS) using ex ante variance rather than with OLS.

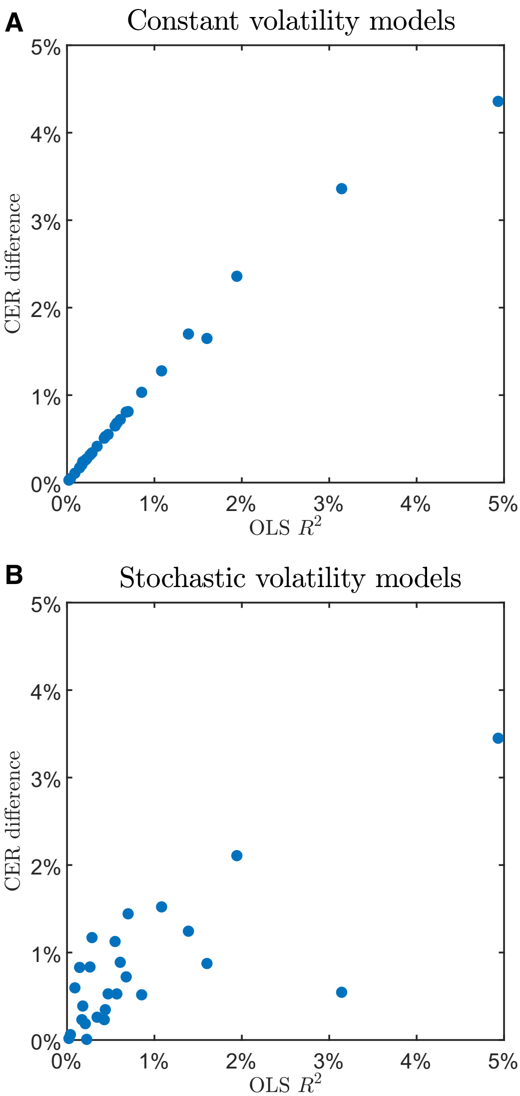

Figure 1. OLS R^2s and in-sample CER gains. The figure shows scatter plots of the OLS R^2 from a predictive regression of log market excess returns on a predictor variable and the CER difference that captures the economic significance of return predictability. The predictive R^2s are from monthly regressions. The 1-month, CV (SV) CER gain in Panel A (Panel B) for a predictor is calculated as the difference between the CERs of the P-CV (P-SV) investor for the optimal strategies under the P-CV and NP-CV (P-SV and NP-SV) models. The CER gains are annualized by multiplying the monthly CER gains by 12.

volatility and the predictor during the portfolio optimization stage. If a given predictor tends to make extreme return forecasts that coincide with periods of pronounced market volatility, the investor tends to moderate her bets and realize smaller CER gains. We find that the statistical effect, the average volatility effect, and the portfolio effect vary in importance across the predictors.

Considering SV is important for determining the in-sample economic significance of predictability evidence. Some predictors have larger utility gains, as is the case for the Treasury bill yield, which has a CER gain of 1.17% per year with SV versus only 0.34% with CV. Alternatively, some predictors are less economically significant, such as the partial least squares aggregated book-to-market ratio of Kelly and Pruitt (2013) with a CER gain of only 0.55% in the SV model versus 3.36% in the CV model. Overall, considering SV increases perceived economic value for eleven of the twenty-six predictors (seven of the fourteen Goyal–Welch predictors and four of the twelve new predictors), whereas benefits fall for the remaining fifteen variables. The most pronounced increases occur for variables from Goyal and Welch (2008) related to interest rate levels. Several of the new predictors, in contrast, tend to take extreme values during high market volatility periods, which lead to lower utility gains. Panel B of Figure 1 plots the association between OLS R^2 and CER gain across all twenty-six variables. The strong relation observable in the CV results is much less pronounced in the SV framework (the Spearman rank correlation is 0.670).

We also consider the in-sample effect of investment horizon. Fama and French (1988) demonstrate stronger statistical significance for multi-period return predictability, which occurs when predictors are highly persistent (Boudoukh, Richardson, and Whitelaw, 2008). We specify Bayesian investors with multi-period horizons to quantify the economic importance of persistence in return forecasts. The CER gain from considering predictability evidence is lower when a given predictor variable is less persistent because expected return quickly converges toward its mean as horizon increases. In addition to this direct channel, the perceived risk of stocks is higher because of uncertainty about future expected returns, particularly for predictors with low persistence and high volatility (e.g., the VRP). We show large effects of horizon on CER gains associated with many of the twenty-six predictors.

Whereas the in-sample design provides a clean setting for demonstrating the economic mechanisms stemming from volatility and horizon, the practical appeal of return predictability rests in the potential for investment strategies based on real-time forecasts. To study this important issue, we implement an out-of-sample, expanding-window approach for each model and predictor. In each month of the investment period, a Bayesian investor estimates a model using historical data and determines the optimal portfolio weight, and we use this solution as the investor's weight in stocks in the next month. After constructing the full time series of out-of-sample weights, we evaluate the portfolio strategies for the models with predictability relative to benchmark models with no predictability. We specifically compute annual out-of-sample CER gains from considering predictability.

Accounting for time-varying volatility greatly improves the potential for investment strategies based on real-time forecasts. In the CV setting, only eleven of the twenty-six predictors produce a positive CER gain. Just six generate a CER gain of at least 0.5% per year and two are statistically significant at the 10% level. In the SV setting, positive CER gains are more common and larger on average. Twenty of the twenty-six predictors (twelve from Goyal and Welch, 2008, and eight from newer studies) produce CER gains. Seventeen of these gains exceed 0.5% per year. Eleven are statistically significant. As an illustration of the potential value of considering time-varying volatility, the Treasury bill yield strategy in the CV framework generates a utility loss with a CER difference of –0.13% per year. In the SV framework, this predictor generates a large, statistically significant CER gain of

3.86% per year. The success of a broad set of predictors in a real-time setting demonstrates the practical relevance of considering time-varying volatility.

Our study fits within the literature investigating return predictability through the lens of Bayesian investors. Kandel and Stambaugh (1996); Stambaugh (1999); Barberis (2000); and Wachter and Warusawitharana (2009, 2015) consider stock market allocations based on return predictability, as in our study.[4] Avramov (2002, 2004) and Tu and Zhou (2010) investigate portfolio choice with multiple assets that have predictable returns. Different from these papers, we incorporate SV into returns and predictors while interpreting the return predictability evidence. There is also an established literature that assesses the economic implications of predictability for investing in mutual funds (e.g., Avramov and Wermers, 2006; Banegas *et al.*, 2013), hedge funds (e.g., Avramov *et al.*, 2011; Avramov, Barras, and Kosowski, 2013), and pension funds (e.g., Christopherson, Ferson, and Glassman, 1998).[5]

Three papers that consider time-varying market volatility in the context of Bayesian investors are more closely related to ours. Shanken and Tamayo (2012) model returns with time-varying volatility and with expected return as a function of time-varying volatility and payout yield. Johannes, Korteweg, and Polson (2014) examine whether real-time investors can benefit from information in payout yield and SV. Relative to these studies, we consider a substantially broader set of return predictors and we find many of the most interesting effects in predictors that are not closely related to payout yield. Pettenuzzo, Timmermann, and Valkanov (2014) introduce SV in market returns in the context of specifying a constraint on the conditional market Sharpe ratio while estimating the predictive return regression. We directly study and quantify the impact of SV on predictability and portfolio choice, and we consider a large set of new predictors in addition to those of Goyal and Welch (2008).

The rest of the paper is organized as follows. Section 2 introduces the Bayesian investor's problem, the models for stock market returns, and our estimation procedures. Section 3 discusses the data. Section 4 develops analytical results on conditional return moments and portfolio choice. Section 5 presents our in-sample results and Section 6 presents our out-of-sample results. Section 7 provides a summary of the main results and Section 8 concludes.

2. Methodology

This section develops our approach to investigating the economic importance of stock market return predictors. Section 2.1 introduces the Bayesian investor's problem. Section 2.2 describes models for returns that either do not or do incorporate return predictability and SV. Section 2.3 discusses estimation and Section 2.4 lays out our approaches to measuring the economic significance of a given return predictor in the in-sample and out-of-sample settings.

4 In a related study, Cremers (2002) compares the posterior views of Bayesian investors with priors that are skeptical and confident about predictability.

5 Like our paper, many of these studies employ a Bayesian framework. See Avramov and Zhou (2010) for a review. Rapach and Zhou (2022) provide a review of the return predictability literature using a frequentist approach.

2.1 Bayesian Investor

We consider a Bayesian investor who chooses optimal allocations between the stock market and a risk-free security. The investor has power utility over wealth at a horizon of k months,

$$U(W_{T+k}) = \frac{W_{T+k}^{1-\gamma}}{1-\gamma}, \tag{2}$$

where γ is the coefficient of relative risk aversion. Wealth at time $T+k$ is given by

$$W_{T+k} = W_T(R_{f,T\to T+k} + \omega_T R_{T\to T+k}), \tag{3}$$

where $R_{f,T\to T+k}$ is the risk-free rate for a k-month horizon, $R_{T\to T+k}$ is the cumulative stock market return in excess of the risk-free rate, and ω_T is a portfolio allocation to stocks that is chosen at time T. We consider investors with $\gamma=5$ and horizons ranging from 1 to 6 months.[6]

Investor i's beliefs about stock market return dynamics are based on a model $\mathcal{M}_i$. The investor maximizes expected utility at time T by choosing an optimal allocation to stocks,

$$\max_{\omega_T} E[U(W_{T+k})|\mathcal{M}_i, D_T], \tag{4}$$

where the conditional expectation is taken with respect to the predictive distribution of excess stock market returns,

$$p(R_{T\to T+k}|\mathcal{M}_i, D_T) = \int p(R_{T\to T+k}|\mathcal{M}_i, \theta, D_T)p(\theta|\mathcal{M}_i, D_T)d\theta, \tag{5}$$

in which θ is the set of parameters in model $\mathcal{M}_i$, D_T denotes the time series of returns and state variables in model $\mathcal{M}_i$, and $p(\theta|\mathcal{M}_i, D_T)$ is the posterior distribution of θ. The predictive distribution of excess returns in Equation (5) accounts for uncertainty about the parameters in the return process, such that the conditional expectation in Equation (4) integrates over this uncertainty.

2.2 Return Process

We study the implications of stock return predictability, SV, and the interaction of these two effects for Bayesian investors' utility. As such, we specify four alternative models featuring (i) no predictability with CV (NP-CV), (ii) predictability with CV (P-CV), (iii) no predictability with SV (NP-SV), and (iv) predictability with SV (P-SV). Given a candidate predictor variable x_t, the processes for the excess stock market return and the state variable are

$$r_{t+1} = \alpha + \beta x_t + \epsilon_{t+1}^r, \tag{6}$$

$$x_{t+1} = \alpha_x + \beta_x x_t + \epsilon_{t+1}^x, \tag{7}$$

where r_{t+1} is the log excess return in month $t+1$.[7] Following much of the return predictability literature, the expected log excess return is specified as a linear function of x_t in

6 Results corresponding to investors with $\gamma=2$ or $\gamma=8$ are available in Online Appendix F, and inferences about return predictors are similar to those in the $\gamma=5$ base case.

7 We assume x_0 is nonstochastic. Given this assumption, the predictive regression coefficient is unbiased. See Stambaugh (1999) and Wachter (2010) for additional discussion.

models P-CV and P-SV, and the predictor variable follows a stationary first-order autoregressive [AR(1)] process. The regression coefficients have a multivariate normal prior distribution centered at 0 with large variance ($10^6 I$). The models with no predictability (NP-CV and NP-SV) have the restriction $\beta = 0$ and all four models have the restriction $-1 < \beta_x < 1$.

The error terms in Equations (6) and (7) are conditionally normally distributed, but the conditional distributions differ across the CV models and the SV models. The errors for the CV models (NP-CV and P-CV) are distributed bivariate normal,

$$\begin{bmatrix} \epsilon^r_{t+1} \\ \epsilon^x_{t+1} \end{bmatrix} \sim N(0, \Sigma), \quad \Sigma = \begin{bmatrix} \sigma^2_r & \sigma_{rx} \\ \sigma_{rx} & \sigma^2_x \end{bmatrix}. \tag{8}$$

The covariance matrix has an inverse-Wishart prior distribution centered at the sample estimate using an empirical Bayes approach and 5 degrees of freedom (the matrix dimension plus 3; see, e.g., Rossi, Allenby, and McCulloch, 2005), such that Σ has a diffuse but proper prior distribution.

The errors for the SV models (NP-SV and P-SV) follow the specification of Primiceri (2005). In particular, ϵ^r_{t+1} and ϵ^x_{t+1} are conditionally normally distributed,

$$\begin{bmatrix} \epsilon^r_{t+1} \\ \epsilon^x_{t+1} \end{bmatrix} \sim N(0, \Sigma_t), \quad \Sigma_t = \begin{bmatrix} \sigma^2_{r,t} & \sigma_{rx,t} \\ \sigma_{rx,t} & \sigma^2_{x,t} \end{bmatrix}. \tag{9}$$

The conditional covariance matrix Σ_t can be decomposed as

$$\Sigma_t = \begin{bmatrix} 1 & 0 \\ a_t & 1 \end{bmatrix} \begin{bmatrix} \sigma^2_{r,t} & 0 \\ 0 & \tilde{\sigma}^2_{x,t} \end{bmatrix} \begin{bmatrix} 1 & a_t \\ 0 & 1 \end{bmatrix}, \tag{10}$$

where $a_t = \sigma_{rx,t}/\sigma^2_{r,t}$, $\tilde{\sigma}^2_{x,t} = \sigma^2_{x,t} - a^2_t \sigma^2_{r,t}$ and the processes for the time-varying parameters are

$$\log(\sigma_{r,t}) = \log(\sigma_{r,t-1}) + \eta^r_t, \quad \eta^r_t \sim N(0, V_r), \tag{11}$$

$$\log(\tilde{\sigma}_{x,t}) = \log(\tilde{\sigma}_{x,t-1}) + \eta^x_t, \quad \eta^x_t \sim N(0, V_x), \tag{12}$$

$$a_t = a_{t-1} + \eta^a_t, \qquad \eta^a_t \sim N(0, V_a). \tag{13}$$

This specification for Σ_t allows for time variation in the conditional volatilities of the return and the state variable as well as time variation in the contemporaneous correlation between the errors. The initial states $\log(\sigma_{r,0})$, $\log(\tilde{\sigma}_{x,0})$, and a_0 have normal prior distributions with mean 0 and variance of 4, and the V_r, V_x, and V_a hyperparameters have inverse-gamma prior distributions with scale parameter 0.1^2 and shape parameter of 2. Additional details on prior parameters for all models are available in Online Appendix A.[8]

8 We investigate the sensitivity of inferences to prior parameters by calculating CER differences between the base case and alternative specifications. Specifically, we consider all P-SV model specifications in which the prior variance of a normal prior distribution or the scale parameter of an inverse-gamma prior distribution either increases or decreases by a factor of 4. The CER differences are economically small in all cases, such that our conclusions are not driven by our choices for prior parameter values. See Online Appendix A for additional details.

2.3 Estimation and Discussion

Each of the four models introduced in Section 2.2 is a restricted Bayesian vector autoregression (BVAR). We implement Markov chain Monte Carlo (MCMC) algorithms to estimate the models. We use a Gibbs sampler to estimate the BVARs for the NP-CV and P-CV models, and we use the Gibbs sampler of Primiceri (2005) and Del Negro and Primiceri (2015) to estimate the NP-SV and P-SV models. See Online Appendix A for additional information.

Before proceeding, two aspects related to the design and estimation of our SV process warrant further discussion. First, some studies specify SV processes with mean reversion, whereas we follow Primiceri (2005) in assuming random walks in Equations (11) and (12) for tractability. In Online Appendix B, we compare our SV estimates for returns from Equation (11) with those from a univariate SV model with mean reversion. The two volatility estimates are nearly indistinguishable with a time-series correlation of 0.997, such that our assumption that log volatility follows a random walk has little impact on our findings.[9] Second, volatility models are often estimated with daily or even intraday data, whereas we use monthly data to estimate the models. The structure of the model in Equations (6) and (7) requires that we use the same frequency of data for returns and predictors and we use monthly predictor data. Further, we show in Online Appendix B that volatility estimates using monthly data provide better forecasts of monthly squared stock returns compared with volatility estimates using daily data. Given that our investors in Section 2.1 have horizons of at least 1 month, the monthly estimates are more relevant in the context of our portfolio choice problem.

2.4 Economic Significance of a Predictor

We examine Bayesian investors' CERs to quantify the economic significance of return predictability in the presence of SV and multi-period investment horizons. Section 2.4.a describes CER calculations for the in-sample design and Section 2.4.b provides the corresponding discussion for the out-of-sample design.

2.4.a. *In-sample CERs*

We evaluate the in-sample significance of information from a predictor variable by comparing the CER for the optimal policy from a model that includes the predictor with the CER that corresponds to the policy that would be optimal under an otherwise similar model without predictability. For example, to discern the effect of return predictability in a setting with SV, we compare the CER for the optimal P-SV policy with the CER that the P-SV investor assigns to the optimal policy for the NP-SV model. In these cases, expected utility is taken with respect to the predictive return distribution from the P-SV model and the comparison of CERs for the P-SV and NP-SV models isolates the economic effect of the return predictability signal in the SV setting.

9 We also compare our SV process with other volatility frameworks from the literature in Online Appendix B. Our SV process identifies the same high- and low-volatility periods as the generalized autoregressive conditional heteroskedasticity (GARCH) model, the exponential GARCH (EGARCH) model, and realized variance (RV) forecasts using lagged RV estimates. The time-series correlations between our estimates and the estimates from these models are 0.880 for GARCH, 0.883 for EGARCH, and 0.877 for RV.

Formally, consider investor i who forms beliefs about the predictive return distribution in Equation (5) using model $\mathcal{M}_i$. Investor i's CER at time T with the optimal policy under model $\mathcal{M}_i$ denoted as $\omega^*_{i,T}$,

$$\text{CER}_i^{is} = [(1-\gamma)W_T^{\gamma-1}E[U(W_T(R_{f,T\to T+k} + \omega^*_{i,T}R_{T\to T+k}))|\mathcal{M}_i, D_T]]^{\frac{1}{1-\gamma}}, \tag{14}$$

can be compared with investor i's CER from adopting the optimal policy $\omega^*_{j,T}$ from an alternative model $\mathcal{M}_j$,

$$\text{CER}_{i,j}^{is} = [(1-\gamma)W_T^{\gamma-1}E[U(W_T(R_{f,T\to T+k} + \omega^*_{j,T}R_{T\to T+k}))|\mathcal{M}_i, D_T]]^{\frac{1}{1-\gamma}}. \tag{15}$$

The CER difference, $\Delta\text{CER}_{i,j}^{is} = \text{CER}_i^{is} - \text{CER}_{i,j}^{is}$, reflects the economic magnitude of the difference between the optimal policies under models $\mathcal{M}_i$ and $\mathcal{M}_j$ from investor i's perspective. Note that these in-sample CER gains do not distinguish which model provides a better statistical fit to the data or objectively better ex ante return forecasts. Rather, these metrics reflect the extent of differences in the predictive return distributions of the models. This method of measuring the economic significance of information is used by Kandel and Stambaugh (1996); Pástor and Stambaugh (2000); and Avramov (2004), among others.

We are particularly interested in the time-series properties of predictors and the corresponding effects on economic significance. Several of the predictor variables take on extreme values, often during periods of high stock market volatility. As such, our in-sample analysis preserves the time-series properties of the data. To evaluate a given predictor, we first estimate each model using the full time series of data to produce a posterior distribution of parameters, $p(\theta|\mathcal{M}_i, D_T)$. We then consider the predictive return distribution in each month of the sample period. We denote months by $\tau = 1, \ldots, T$ in this stage of the analysis to differentiate from the $t = 1, \ldots, T$ notation in the estimation stage. The predictive distribution for model $\mathcal{M}_i$ in each month τ is based on the posterior draws of parameters $[p(\theta|\mathcal{M}_i, D_T)]$ and the value of the predictor variable in that month (x_τ). In the models with SV, we maintain the relation between x_τ and Σ_τ that is estimated from the data, which is important for our goal of investigating interactions between return predictability and SV. Our analysis thus takes the perspective of a hypothetical investor whose beliefs about return predictability and volatility mirror the full-sample posterior distribution (reflecting the in-sample design). This investor optimizes in each month of the sample based on the observed predictor variable (preserving the time-series properties of the data).

We estimate the in-sample CER difference between models $\mathcal{M}_i$ and $\mathcal{M}_j$ using the predictive distributions for each month τ of the sample. For each model and month, we calculate the optimal weight in stocks for the investor described in Section 2.1. We then calculate the time-series averages of the expected utilities that investor i gains from using the optimal portfolio policies from models $\mathcal{M}_i$ and $\mathcal{M}_j$. We calculate CER_i^{is} and $\text{CER}_{i,j}^{\text{is}}$ according to Equations (14) and (15) and we annualize by multiplying k-horizon CERs by $12/k$ and compute the CER difference $\Delta\text{CER}_{i,j}^{\text{is}}$. Additional details about these calculations are provided in Online Appendix A.

2.4.b. *Out-of-sample CERs*

We implement an out-of-sample portfolio design to complement the in-sample approach. The in-sample CERs described in Section 2.4.a correspond to investors who learn about predictability from the full sample of data, which would not have been available to investors in real time. Investors in our out-of-sample approach, in contrast, must learn solely from historical data when determining their optimal portfolio weights.

We adopt an expanding-window design for our out-of-sample tests. For a given model $\mathcal{M}_i$ and an estimation window of length τ, we estimate the model using data on returns $(r_1, \ldots, r_\tau)$ and predictors $(x_0, \ldots, x_{\tau-1})$ that were available at the end of month τ of the sample. We then produce the horizon-k predictive distribution of $R_{\tau \to \tau+k}$ (i.e., the cumulative return earned in months $\tau+1$ to $\tau+k$) using the predictor value x_τ and posterior $p(\theta|\mathcal{M}_i, D_\tau)$. The investor described in Section 2.1 maximizes expected utility with the optimal portfolio weight $\omega^*_{i,\tau}$. The minimum estimation window length is 240 months, such that $\tau = 240, \ldots, T-1$. The expanding-window approach therefore produces a time series of out-of-sample portfolio weights $\omega^*_{i,240}, \ldots, \omega^*_{i,T-1}$.

Given the optimal weights for a given model $\mathcal{M}_i$, we evaluate the realized returns on the strategy portfolio. With a 1-month horizon, strategy i's portfolio return in month $\tau+1$ is

$$R_{i,\tau+1} = R_{f,\tau+1} + \omega^*_{i,\tau} R_{\tau+1}, \tag{16}$$

where $R_{f,\tau+1}$ is the risk-free rate and $R_{\tau+1}$ is the excess market return. We calculate the power utility investor's annualized out-of-sample CER for each model and predictor given the time series of realized strategy returns. Of note, these out-of-sample CERs do not directly depend on the investor's beliefs about model parameters or the predictive return distribution, which differentiates their interpretation from the in-sample CERs that depend on expectations taken under the investor's predictive distribution. The out-of-sample CERs, rather, condition on the time-series realization of out-of-sample strategy returns, and they represent utility gains that real-time investors could have achieved by investing in the strategies (assuming no trading costs and the ability to borrow at the risk-free rate).

We compare out-of-sample CERs across the models with and without predictability to assess the real-time value of a given predictor. To assess the statistical significance of the CER difference between models $\mathcal{M}_i$ and $\mathcal{M}_j$ (i.e., $\Delta\text{CER}^{\text{os}}_{i,j}$), we design a bootstrap approach that resamples the paired out-of-sample strategy return realizations (i.e., $R_{i,\tau}$ and $R_{j,\tau}$). We form bootstrap samples of $T-240$ months of strategy returns to match the out-of-sample period length and calculate the CER difference for each bootstrap sample. We then compute the percentage of bootstrap CER differences that are negative to determine a bootstrap P-value that corresponds to the one-sided test of the null hypothesis that the gain is less than zero. Additional details on out-of-sample estimation, CER calculations, and bootstrap design are available in Online Appendix A.

3. Data

Our empirical tests focus on forecasting log excess stock market returns using a variety of predictor variables. The market portfolio is the Center for Research in Securities Prices (CRSP) value-weighted index. We collect monthly time-series data on the excess market return and the risk-free rate from Kenneth French's website.[10] The log excess market return is the log return on the CRSP index less the log return on the risk-free asset.

We consider a wide range of forecasting variables from prior literature. We examine stock return predictability at a monthly horizon, so we restrict the sample to predictors that

10 See http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/. We thank Kenneth French for making these data available.

are available at a monthly frequency. We also require that each predictor variable has data availability through December 2017. We provide a brief overview of the predictors below. Full details on variable definitions, data sources, and construction methods are available in Online Appendix C.

We start with the fourteen monthly predictor variables from Goyal and Welch (2008). This set of predictors includes the dividend-price ratio (DP), the dividend yield (DY), the earnings-price ratio (EP), the dividend-earnings ratio (DE), stock market variance (SVAR), the book-to-market ratio (BM), net equity expansion (NTIS), the Treasury bill yield (TBL), the long-term Treasury bond yield (LTY), the long-term Treasury bond return (LTR), the term spread (TMS), the default yield spread (DFY), the default return spread (DFR), and inflation (INFL). This group of forecasting variables is widely used in the literature on stock market return predictability.

We augment the Goyal–Welch predictors with a second group of twelve forecasting variables introduced in more recent literature. We specifically search articles appearing in top finance journals subsequent to the publication of Goyal and Welch (2008) and identify predictors with data availability at a monthly frequency. This group includes Kelly and Pruitt's (2013) partial least squares aggregated book-to-market ratio (PLS), Chen and Zhang's (2011) employment growth (EG), Cooper and Priestly's (2009) output gap (GAP), Jones and Tuzel's (2013) new orders-to-shipments of durable goods (NOS), Li and Yu's (2012) nearness to Dow historical high (DOW), Kelly and Jiang's (2014) tail risk (TAIL), Pollet and Wilson's (2010) average correlation (COR), Rapach, Ringgenberg, and Zhou's (2016) short interest index (SII), Huang and Kilic's (2019) gold-to-platinum price ratio (GP), Driesprong, Jacobsen, and Maat's (2008) oil price change (OIL), Bollerslev, Tauchen, and Zhou's (2009) VRP, and Bollerslev, Todorov, and Xu's (2015) left jump tail variation (LJV).

Table I reports summary statistics for the twenty-six predictor variables. Panel A presents the sample period start date, mean, standard deviation, skewness, and kurtosis for each of the Goyal–Welch predictors and Panel B shows the corresponding statistics for the new predictors. The times series for the Goyal–Welch variables in Panel A and the PLS variable in Panel B begin in January 1927. Data for the other new predictors cover shorter sample periods.

Table I highlights distributional properties of the predictors that are relevant for our subsequent analysis. In particular, the empirical distributions for several of the forecasting variables are highly non-normal, as indicated by the skewness and kurtosis statistics. In the standard univariate predictive regression setting, non-normal forecasting variables imply that conditional expected stock returns are also non-normal and tend to take on extreme values in a few sample months. Deviations from normality are acute for several of the new predictor variables. Both PLS and VRP, for example, are highly negatively skewed. These variables are positive return predictors, such that negative skewness translates into extreme negative expected return forecasts for months in which PLS and VRP take on their lowest values. The PLS and VRP predictors also exhibit fat tails, with kurtosis measures of 32.62 and 56.99, respectively.

Table II presents results from standard univariate predictive regressions of excess stock market returns [i.e., Equation (6)] and predictor variables [i.e., Equation (7)]. For each regression, the table reports the OLS estimate of the slope coefficient, its corresponding two-tailed P-value based on the OLS standard error, and the regression R^2.

The state variable regression estimates in Table II reveal substantial variation in measures of persistence across the predictors. Fifteen of the twenty-six forecasting variables have

Table I. Summary statistics for predictor variables

The table reports summary statistics for stock market return predictor variables. Panel A shows summary statistics for the Goyal–Welch predictors and Panel B displays summary statistics for the new predictors. All predictor variables are monthly such that the summary statistics reflect monthly values. The sample period end date for each predictor is December 2017.

Predictor	Sample start	Mean	Standard deviation	Skewness	Kurtosis
Panel A: Goyal and Welch (2008) predictors					
DP	1927:01	−3.37	0.46	−0.22	2.66
DY	1927:01	−3.32	0.45	−0.46	2.73
EP	1927:01	−2.74	0.42	−0.60	5.62
DE	1927:01	−0.64	0.33	1.52	9.03
SVAR	1927:01	0.00	0.01	5.79	46.65
BM	1927:01	0.57	0.27	0.78	4.46
NTIS	1927:01	0.02	0.03	1.65	11.25
TBL	1927:01	0.03	0.03	1.08	4.28
LTY	1927:01	0.05	0.03	1.08	3.60
LTR	1927:01	0.00	0.02	0.59	7.69
TMS	1927:01	0.02	0.01	−0.29	3.17
DFY	1927:01	0.01	0.01	2.48	11.87
DFR	1927:01	0.00	0.01	−0.39	10.79
INFL	1927:01	0.00	0.01	1.08	16.82
Panel B: New predictors					
PLS	1927:01	−0.71	0.34	−4.68	32.62
EG	1939:05	0.01	0.01	−0.20	11.54
GAP	1947:12	0.00	0.07	−0.03	1.98
NOS	1958:02	0.01	0.04	−0.01	4.72
DOW	1960:01	0.90	0.10	−1.11	3.93
TAIL	1963:01	0.00	1.00	−0.49	2.79
COR	1963:03	0.26	0.11	0.86	4.36
SII	1973:01	0.00	0.25	0.38	2.97
GP	1975:01	−0.20	0.28	−0.57	2.51
OIL	1983:04	0.00	0.09	−0.21	5.33
VRP	1990:01	16.20	20.40	−3.72	56.99
LJV	1996:06	0.00	0.00	2.96	14.43

monthly autocorrelation coefficients that exceed 0.95. Five predictors, in contrast, have autocorrelation coefficients below 0.50 in magnitude. Autocorrelation is an important summary measure because more persistent predictors tend to be more influential in settings that require multi-period return forecasts.

For the return regressions in Panel A of Table II that use the Goyal–Welch predictors, the statistical evidence in favor of predictability is modest. Only six of the fourteen variables are statistically significant return predictors at the 10% level and none is significant at the 1% level. The monthly R^2 values tend to be small, ranging from 0.000 to 0.006.

Table II. Predictive regression coefficients from OLS

The table reports predictive regression coefficients, associated *P*-values against the null of no predictability, and regression R^2s from OLS regressions. Panel A shows predictive regression results for the Goyal–Welch predictors and Panel B displays results for the new predictors. The predictive regressions are monthly log stock market excess returns on lagged predictor variables and monthly predictor variables on lagged predictor variables.

	Stock market return			State variable		
Predictor	β	*P*-value	R^2	β_x	*P*-value	R^2
Panel A: Goyal and Welch (2008) predictors						
DP	0.005	0.161	0.002	0.993	0.000	0.986
DY	0.008	0.029	0.004	0.993	0.000	0.986
EP	0.008	0.052	0.003	0.987	0.000	0.974
DE	–0.002	0.623	0.000	0.991	0.000	0.983
SVAR	–0.383	0.173	0.002	0.633	0.000	0.400
BM	0.013	0.031	0.004	0.986	0.000	0.972
NTIS	–0.157	0.012	0.006	0.981	0.000	0.962
TBL	–0.092	0.077	0.003	0.993	0.000	0.986
LTY	–0.073	0.210	0.001	0.996	0.000	0.992
LTR	0.113	0.090	0.003	0.043	0.152	0.002
TMS	0.188	0.131	0.002	0.961	0.000	0.924
DFY	0.152	0.516	0.000	0.975	0.000	0.951
DFR	0.187	0.117	0.002	–0.121	0.000	0.015
INFL	–0.300	0.323	0.001	0.481	0.000	0.231
Panel B: New predictors						
PLS	0.028	0.000	0.031	0.962	0.000	0.925
EG	–0.388	0.010	0.007	0.887	0.000	0.787
GAP	–0.091	0.000	0.019	0.989	0.000	0.979
NOS	–0.119	0.005	0.011	0.655	0.000	0.435
DOW	–0.032	0.070	0.005	0.938	0.000	0.879
TAIL	0.004	0.017	0.009	0.816	0.000	0.667
COR	0.030	0.045	0.006	0.900	0.000	0.809
SII	–0.015	0.056	0.007	0.972	0.000	0.941
GP	0.019	0.007	0.014	0.990	0.000	0.975
OIL	–0.034	0.130	0.005	0.172	0.000	0.030
VRP	0.046	0.000	0.049	0.276	0.000	0.076
LJV	4.274	0.041	0.016	0.959	0.000	0.922

Panel B of Table II shows that the new predictors generate much stronger statistical support for return predictability. Eleven of the twelve variables are statistically significant at the 10% level and six remain significant at the 1% level. The regression R^2s are also more impressive than those in Panel A, with R^2 values as high as 0.031 for PLS and 0.049 for VRP. Based on the analysis in Kandel and Stambaugh (1996), these results suggest that many of the new predictors should be of considerable economic value to investors making asset allocation decisions. At the same time, however, the non-normal distributions for these variables and their other time-series properties may limit their value in portfolio applications.[11]

4. Analytical Results and Predictive Return Distributions

Our framework in Section 2 is based on a Bayesian investor who optimizes power utility given a predictive distribution of returns. In this section, we develop analytical results to better understand the sources of economic value from return predictability. To best highlight the economic links between predictor properties and portfolio choice, we focus in this section on the in-sample design in which the investor conditions on the full-sample posterior distribution of model parameters. The analysis draws on the fact that the portfolio-choice decision introduced in Section 2.1 is well described by its dependence on conditional return moments. In particular, the optimal weight in stocks chosen by the power utility investor using model $\mathcal{M}_i$ is closely approximated by (e.g., Kandel and Stambaugh, 1996)

$$\omega_{i,\tau}^* \approx \frac{\mu_{i,\tau}}{\gamma\sigma_{i,\tau}^2} + \frac{1}{2\gamma}, \tag{17}$$

where $\mu_{i,\tau}$ and $\sigma_{i,\tau}^2$ are defined to be the conditional mean and variance, respectively, of the time-τ predictive distribution of log excess returns. Equation (17) illustrates that considering return predictability will generate time variation in optimal portfolio weights through variation in $\mu_{i,\tau}$, all else equal. In addition, in models with time-varying volatility, the effect of return predictability on portfolio choice is dependent on the interaction between $\mu_{i,\tau}$ and $\sigma_{i,\tau}^2$. The investor's reaction to the information content of $\mu_{i,\tau}$ will be dampened (amplified) when $\sigma_{i,\tau}^2$ is relatively high (low). Section 4.1 considers the behavior of $\mu_{i,\tau}$ and $\sigma_{i,\tau}^2$ given the structure of the VARs in Section 2.2. Section 4.2 formally links conditional return moments to in-sample CER gains and develops our approach to quantifying the economic effects of variation in (and interactions between) conditional moments.

4.1 Predictive Return Moments

According to Equation (17), the optimal portfolio weight relates to the mean and variance of the predictive return distribution. We proceed to derive these conditional moments as functions of parameters and state variables. As described in Section 2.4.a, we produce a predictive return distribution for each month τ that conditions on the full-sample posterior distribution and month-τ state variables, such that we calculate conditional moments corresponding to each month τ and horizon k. We specialize in this section to a single-period horizon and corresponding results for multi-period horizons are available in Online Appendix D.

11 It is also important to consider the correlation between the estimated innovations to returns and predictor variables. For a positive return predictor [i.e., one with $\hat{\beta} > 0$ in Equation (6)] in a setting with constant volatility, for example, the innovations in Equations (6) and (7) would need to be negatively correlated for increases in expected future returns to be associated with low current returns (Campbell, 1991). The required relation is more complex in a setting with SV, but a predictor with a large positive correlation of innovations could raise some concern. To examine this issue, we compute $\text{sign}(\hat{\beta}) \times \text{corr}(\hat{\epsilon}_{t+1}^r, \hat{\epsilon}_{t+1}^x)$ for each predictor. The relatively large positive values of this quantity for TAIL (0.47), SVAR (0.29), and DFR (0.17) signal problems with the plausibility of the return processes implied by these three variables.

The conditional mean of the predictive distribution of log excess returns is

$$\mu_{i,\tau} = E(\alpha + \beta x_\tau | \mathcal{M}_i, D_T). \tag{18}$$

The quantities $E(\alpha|\mathcal{M}_i, D_T)$ and $E(\beta|\mathcal{M}_i, D_T)$ are posterior means of the α and β parameters from the predictive regression in Equation (6). Recall that β is restricted to equal zero for the NP-CV and NP-SV models, such that the conditional mean is constant.

For the models with predictability, inferences about α and β can differ substantially across the P-CV and P-SV models. Estimation in the CV framework is similar in spirit to OLS in which periods are equally weighted, whereas the SV model is similar to WLS in that it weights information from each period based on conditional variance. For a given predictor, the relative variation in the conditional mean implied by the P-CV and P-SV models is dependent on the relative magnitude of β.

To demonstrate the differences in β estimates across models, Figure 2 shows quantiles of β posterior distributions for each predictor under the P-CV and P-SV models. Panel A shows results for the Goyal–Welch predictors and Panel B displays posteriors for the new predictors. For each posterior, we plot the posterior median, interquartile range, and 95% credible interval. The posteriors for many of the predictors are similar across models, but there are several with important differences. For instance, there is greater than 75% posterior probability that returns are more predictable under the P-SV model compared with the P-CV model using the TBL, LTY, and INFL variables, whereas returns are less predictable under the P-SV model with more than 75% posterior probability for the DY, BM, DFR, PLS, and GP predictors.

The predictive variance of single-period log returns has two components,

$$\sigma^2_{i,\tau} = E(\sigma^2_{r,\tau}|\mathcal{M}_i, D_T) + \text{Var}(\alpha + \beta x_\tau|\mathcal{M}_i, D_T). \tag{19}$$

The first term corresponds to the posterior mean of the conditional variance of the single-period return shock. The second term measures estimation risk from posterior uncertainty about the conditional mean. This feature of the asset allocation problem for a Bayesian investor is introduced by Klein and Bawa (1976). Parameter uncertainty increases the risk of stocks from the investor's perspective relative to an environment with known parameters. All else equal, the estimation risk term will be smaller when the model parameters are more precisely estimated. Ex ante, we may expect β posteriors to be somewhat tighter under the SV framework, analogous to the greater efficiency of WLS versus OLS, which would lead to a smaller contribution of estimation risk to predictive variance. Most of the posteriors in Figure 2 support this prediction. In addition, the value of the predictor variable x_τ affects estimation risk for the P-CV and P-SV models. If x_τ takes on an extreme value, estimation risk is amplified because the magnitude of $\text{Var}(\alpha + \beta x_\tau|\mathcal{M}_i, D_T)$ is dependent on x^2_τ.

We calculate $\mu_{i,\tau}$ and $\sigma^2_{i,\tau}$ for each month τ based on x_τ and the full-sample posterior parameters (including Σ_τ). Table III shows the time-series mean and standard deviation of $\mu_{i,\tau}$ and $\sigma_{i,\tau}$ for each model specification. Panel A reports results for the Goyal–Welch predictors and Panel B shows statistics for the new predictors. All statistics are reported in percentage per month.

There are several notable results in Panel A of Table III. First, given that the NP-CV models share the same sample period and do not incorporate predictability, the model moments are virtually indistinguishable across predictors. Second, within each volatility framework, the volatility estimates are not strongly influenced by predictability or a

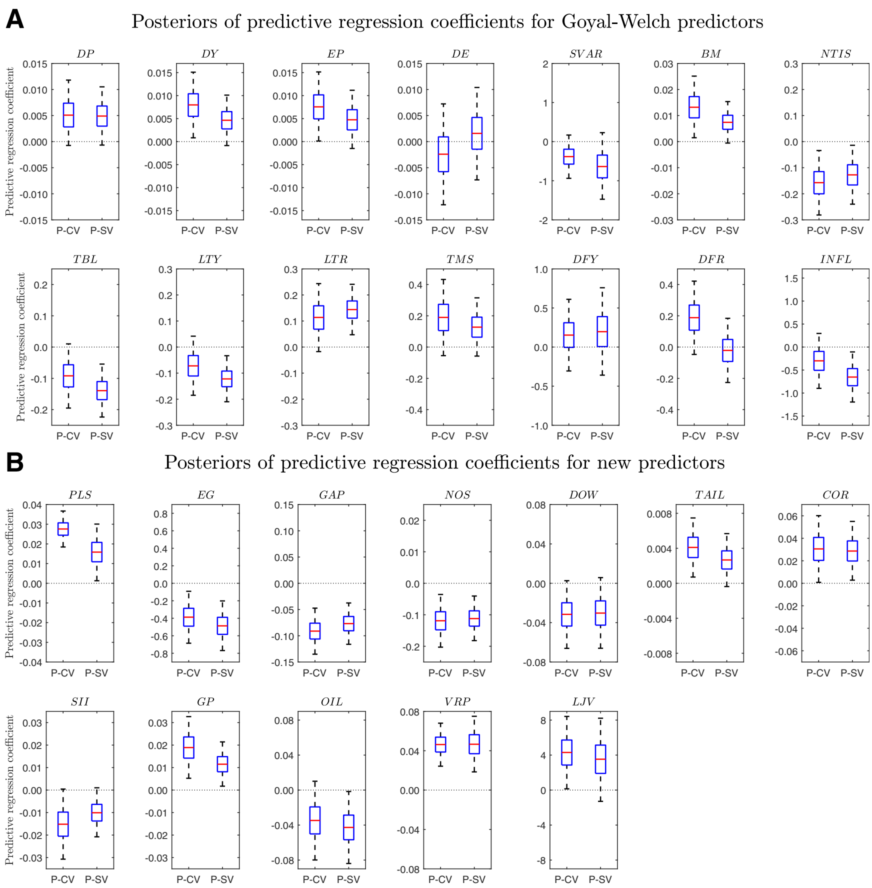

Figure 2. Posteriors of predictive regression coefficients in the CV and SV models. The figure shows a box-and-whiskers plot for the posterior distribution of the predictive regression coefficient for each model and predictor combination. Panel A shows posteriors for the Goyal–Welch predictors and Panel B shows posteriors for the new predictors. Each predictor has posteriors for the P-CV and P-SV models. In each box-and-whiskers plot, the center line shows the posterior median, the box represents a 50% credible interval, and the whiskers span a 95% credible interval.

particular predictor. That is, volatility estimates for a given model are similar across predictors and volatility estimates for a given predictor are similar for the NP-CV and P-CV cases as well as for the NP-SV and P-SV cases. Third, the predictive standard deviation $\sigma_{i,\tau}$ varies over time for the P-CV models (i.e., the time-series standard deviation of $\sigma_{i,\tau}$ is positive) because the estimation risk component of Equation (19) is dependent on x_τ, but the magnitude of this variation is small.

Fourth, and most important, the time-series standard deviations of $\mu_{i,\tau}$ for the P-CV and P-SV models indicate the degree of return predictability in each model, and Equation (17) shows that time variation in $\mu_{i,\tau}$ translates directly into variation in the optimal weight $\omega^*_{i,\tau}$. In the CV case, there is virtually a perfect correlation between the magnitude of variation in $\mu_{i,\tau}$ and the OLS R^2 from a predictive regression. The most volatile $\mu_{i,\tau}$ among the P-CV models in Panel A of Table III, for example, belongs to NTIS, which also has the highest

Table III. Time-series means and standard deviations of conditional return moments

The table reports the time-series mean and standard deviation of the conditional mean (i.e., $\mu_{i,\tau}$) and standard deviation (i.e., $\sigma_{i,\tau}$) of the predictive return distribution from the NP-CV, P-CV, NP-SV, and P-SV models. The conditional moments are monthly figures and all statistics are reported as percentages.

	NP-CV				P-CV				NP-SV				P-SV			
	$\mu_{i,\tau}$		$\sigma_{i,\tau}$		$\mu_{i,\tau}$		$\sigma_{i,\tau}$		$\mu_{i,\tau}$		$\sigma_{i,\tau}$		$\mu_{i,\tau}$		$\sigma_{i,\tau}$	
Predictor	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Panel A: Goyal and Welch (2008) predictors																
DP	0.46	0.00	5.10	0.00	0.46	0.22	5.10	0.00	0.87	0.00	4.55	2.19	0.92	0.20	4.57	2.20
DY	0.46	0.00	5.10	0.00	0.46	0.32	5.09	0.00	0.81	0.00	4.55	2.18	0.83	0.19	4.55	2.16
EP	0.46	0.00	5.10	0.00	0.46	0.28	5.10	0.01	0.81	0.00	4.55	2.18	0.82	0.18	4.56	2.16
DE	0.46	0.00	5.10	0.00	0.46	0.07	5.10	0.01	0.81	0.00	4.55	2.18	0.82	0.05	4.57	2.19
SVAR	0.46	0.00	5.10	0.00	0.46	0.20	5.10	0.02	0.81	0.00	4.55	2.18	0.72	0.33	4.56	2.18
BM	0.46	0.00	5.10	0.00	0.46	0.32	5.09	0.01	0.85	0.00	4.55	2.19	0.90	0.18	4.57	2.19
NTIS	0.46	0.00	5.10	0.00	0.46	0.36	5.09	0.01	0.81	0.00	4.55	2.18	0.77	0.29	4.54	2.14
TBL	0.46	0.00	5.10	0.00	0.46	0.26	5.10	0.00	0.81	0.00	4.55	2.18	0.78	0.39	4.55	2.21
LTY	0.46	0.00	5.10	0.00	0.46	0.18	5.10	0.00	0.83	0.00	4.55	2.18	0.81	0.31	4.55	2.21
LTR	0.46	0.00	5.10	0.00	0.46	0.25	5.10	0.01	0.78	0.00	4.55	2.17	0.78	0.32	4.54	2.17
TMS	0.46	0.00	5.10	0.00	0.46	0.22	5.10	0.00	0.81	0.00	4.55	2.18	0.81	0.15	4.55	2.17
DFY	0.46	0.00	5.10	0.00	0.46	0.10	5.10	0.01	0.81	0.00	4.55	2.18	0.85	0.12	4.57	2.20
DFR	0.46	0.00	5.10	0.00	0.46	0.23	5.10	0.01	0.81	0.00	4.55	2.18	0.81	0.03	4.56	2.18
INFL	0.46	0.00	5.10	0.00	0.46	0.14	5.10	0.01	0.81	0.00	4.55	2.18	0.80	0.31	4.56	2.20
Panel B: New predictors																
PLS	0.46	0.00	5.10	0.00	0.46	0.85	5.03	0.05	0.83	0.00	4.55	2.19	0.70	0.49	4.54	2.09
EG	0.52	0.00	4.14	0.00	0.52	0.33	4.13	0.01	0.78	0.00	3.97	1.16	0.79	0.41	3.96	1.16
GAP	0.50	0.00	4.07	0.00	0.50	0.54	4.04	0.01	0.78	0.00	3.90	1.12	0.75	0.45	3.89	1.10
NOS	0.42	0.00	4.17	0.00	0.42	0.41	4.15	0.01	0.74	0.00	3.98	1.21	0.76	0.38	3.97	1.19
DOW	0.40	0.00	4.20	0.00	0.40	0.27	4.20	0.01	0.71	0.00	4.02	1.21	0.79	0.26	4.03	1.21
TAIL	0.39	0.00	4.21	0.00	0.39	0.37	4.20	0.01	0.71	0.00	4.02	1.23	0.66	0.24	4.01	1.19
COR	0.39	0.00	4.22	0.00	0.39	0.31	4.21	0.01	0.71	0.00	4.02	1.23	0.78	0.29	4.02	1.22
SII	0.41	0.00	4.36	0.00	0.41	0.34	4.36	0.01	0.75	0.00	4.17	1.30	0.73	0.22	4.17	1.28
GP	0.52	0.00	4.22	0.00	0.52	0.47	4.21	0.01	0.80	0.00	4.04	1.20	0.75	0.29	4.04	1.17
OIL	0.51	0.00	4.20	0.00	0.51	0.29	4.20	0.01	0.83	0.00	3.96	1.32	0.85	0.36	3.96	1.33
VRP	0.53	0.00	4.06	0.00	0.53	0.85	3.98	0.10	0.90	0.00	3.85	1.41	1.05	0.86	3.81	1.37
LJV	0.48	0.00	4.32	0.00	0.48	0.52	4.30	0.05	0.90	0.00	4.11	1.46	1.03	0.42	4.12	1.46

predictive regression R^2 in Panel A of Table II. In line with Cochrane (2008, 2011), the predictive regressions produce economically meaningful time variation in expected returns even though the R^2s in Table II are small. We can also compare estimates across the P-CV and P-SV models because both models for a given predictor use the same time series of the predictor variable. The relative variation in $\mu_{i,\tau}$ across models depends on the relative magnitudes of the β estimates that are shown in Figure 2.

Panel B of Table III reports results for the new predictors. Recall that each predictor in Panel B has a unique sample starting date, such that the means of $\mu_{i,\tau}$ and $\sigma_{i,\tau}$ differ substantially across predictors even for the NP-CV model because of the differing sample periods.

Comparing the statistics in Panel B with those in Panel A, the magnitude of time variation in $\mu_{i,\tau}$ tends to be larger for the new predictors, in line with their higher predictive regression R^2s in Table II. This pattern suggests that optimal weights using the new predictors are likely to be more volatile according to Equation (17). The results also indicate that inferences differ substantially across the P-CV and P-SV models. For instance, the time-series standard deviation of $\mu_{i,\tau}$ for the PLS predictor is 0.85% under CV versus only 0.49% with SV.

4.2 Utility and Return Moments

We now demonstrate the relation between the conditional predictive distribution moments and the utility consequences of predictability. As we show in Online Appendix D, the in-sample CER difference between models $\mathcal{M}_i$ and $\mathcal{M}_j$ from the perspective of investor i, $\Delta\text{CER}_{i,j}^{\text{is}}$, is highly dependent on the differences in optimal weights under the two models. Intuitively, investor i perceives little economic loss in switching from $\mathcal{M}_i$ to $\mathcal{M}_j$ if the two models produce similar portfolio implications, whereas large differences between optimal weights could lead to a substantial expected utility loss if the investor were to adopt the suboptimal policy. Formally, the time-τ CER difference between models $\mathcal{M}_i$ and $\mathcal{M}_j$ is proportional to the squared difference between their optimal weights, $(\omega_{i,\tau}^* - \omega_{j,\tau}^*)^2$, up to a second-order Taylor approximation.[12]

Based on this relation and the weight approximation from Equation (17), the average squared weight difference,

$$\text{Average squared weight difference} = \frac{1}{T}\sum_{\tau=1}^{T}\left(\frac{\mu_{i,\tau}}{\gamma\sigma_{i,\tau}^2} - \frac{\mu_{j,\tau}}{\gamma\sigma_{j,\tau}^2}\right)^2, \tag{20}$$

is a moment-based calculation that is informative about the in-sample CER difference between models $\mathcal{M}_i$ and $\mathcal{M}_j$. In our empirical applications in Section 5, we calculate CER differences for the P-CV and NP-CV models in the CV setting and the P-SV and NP-SV models in the SV setting. As such, model $\mathcal{M}_i$ incorporates return predictability so $\mu_{i,\tau}$ is time varying, whereas $\mu_{j,\tau}$ is constant in each case. We also showed in Section 4.1 that volatility estimates are quite similar across models with and without return predictability, such that $\sigma_{i,\tau}^2 \approx \sigma_{j,\tau}^2$ for each CER difference we consider.

Equation (20) encapsulates the effects of time variation in conditional moments. To better understand the underlying mechanisms, we decompose the average squared weight difference into two components. We first introduce a mean component that reflects the difference in the conditional means from models $\mathcal{M}_i$ and $\mathcal{M}_j$. The average squared weight difference that is attributable to the mean component is

$$\text{Mean component} = \left(\frac{\overline{\sigma^{-2}}}{\gamma}\right)^2\left(\frac{1}{T}\sum_{\tau=1}^{T}(\mu_{i,\tau} - \mu_{j,\tau})^2\right), \tag{21}$$

12 Under mean–variance utility, the CER difference is exactly proportional to the squared weight difference. Under more general utility functions (including power utility), the degree of approximation error depends on higher-order derivatives of CER with respect to weight. Across the predictors in our empirical analysis, we observe Spearman rank correlations between CER differences and average squared weight differences of 0.997 for constant volatility and 0.946 for stochastic volatility, indicating that the approximation is informative in our setting. See Online Appendix D for additional information.

where $\overline{\sigma^{-2}}$ is the average inverse of conditional variance from model $\mathcal{M}_i$ (i.e., $\overline{\sigma^{-2}} = \frac{1}{T}\sum_{\tau=1}^{T} \sigma_{i,\tau}^{-2}$).[13] Given that $\mu_{j,\tau}$ is constant for each model comparison, the $\frac{1}{T}\sum_{\tau=1}^{T} (\mu_{i,\tau} - \mu_{j,\tau})^2$ term captures the degree of variation in the conditional mean implied by model $\mathcal{M}_i$. The mean component is, thus, informative about the strength of the return predictability relation and its utility consequences.

The average squared weight difference is the sum of the mean component and a residual component,

$$\text{Residual component} = \frac{1}{T}\sum_{\tau=1}^{T}\left(\frac{\mu_{i,\tau}}{\gamma\sigma_{i,\tau}^2} - \frac{\mu_{j,\tau}}{\gamma\sigma_{j,\tau}^2}\right)^2 - \left(\frac{\overline{\sigma^{-2}}}{\gamma}\right)^2\left(\frac{1}{T}\sum_{\tau=1}^{T}(\mu_{i,\tau} - \mu_{j,\tau})^2\right). \tag{22}$$

In the CV framework, the residual component is relatively small. In the SV setting, a large residual can arise from interactions between $\mu_{i,\tau}$ and $\sigma_{i,\tau}^2$. On the one hand, if a given predictor tends to make extreme return forecasts primarily during periods of high volatility, these predictions will have a muted impact on portfolio choice such that the residual component is negative and the average squared weight difference is smaller than the mean component. On the other hand, the residual component is positive if a predictor tends to produce more extreme forecasts in low-volatility periods, as the investor can react more aggressively such that the average squared weight difference is larger than the mean component. The sign and magnitude of the residual component are, therefore, informative about the direction and strength of interactions between $\mu_{i,\tau}$ and $\sigma_{i,\tau}^2$. In the next section, we use Equations (20)–(22) to better understand in-sample CER gains.

5. In-Sample Results

In this section, we examine the market return predictors from the perspective of Bayesian investors who believe in one of the models introduced in Section 2. Section 5.1 studies the specific example of the dividend-price predictor under CV and a single-period horizon to illustrate our approach to assessing the economic significance of a predictor. Section 5.2 details results for the broad set of predictors under CV and SV in a single-period setting. Section 5.3 discusses the economic significance of return predictability with multi-period horizons.

5.1 Example: Dividend-Price Ratio

We begin our analysis by focusing on the dividend-price ratio (DP) predictor in a setting that closely matches that of Kandel and Stambaugh (1996). As shown in Table II, DP qualifies as a statistically weak predictor with a P-value of 0.16 for the OLS predictability coefficient and a monthly predictive regression R^2 of 0.002. Following Kandel and Stambaugh (1996), we examine DP from the perspective of an investor who believes that volatility is constant.

Figure 3 shows properties of the predictive return distribution and optimal portfolio weights in the CV framework. The top panel presents quantiles of the predictive return distribution from the P-CV model. The model produces a predictive return distribution for each month in the sample period conditional on the value of DP, and the figure plots the median (solid line) and 25th and 75th percentiles (dashed lines) of the distribution. The

13 See Online Appendix D for additional details.

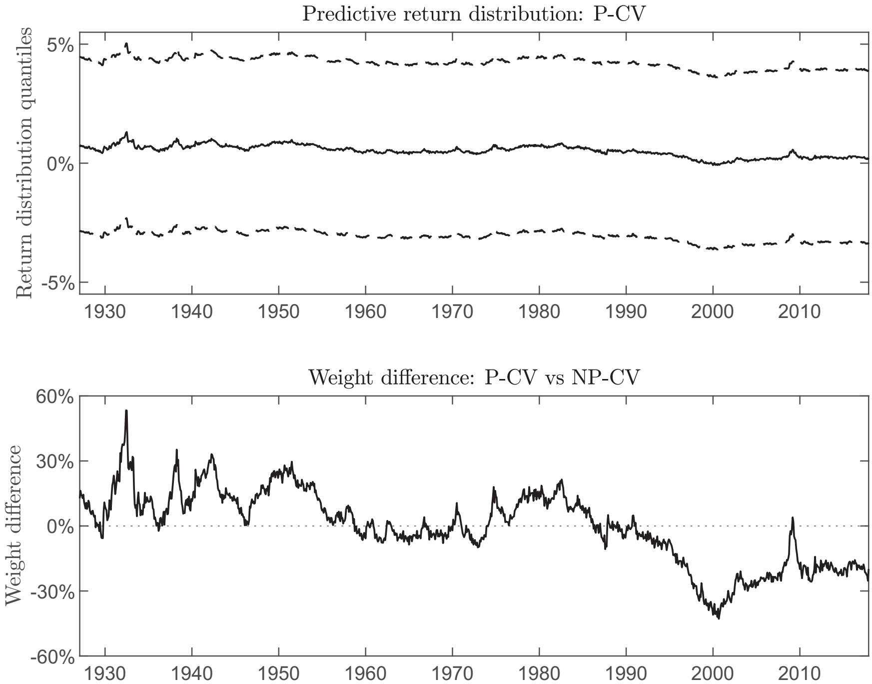

Figure 3. Predictive return distribution quantiles and effects of predictability on portfolio weight for the DP predictor. The figure shows quantiles of the predictive return distributions and weight differences for the CV models in which DP is the market return predictor variable. The top panel shows quantiles of the predictive return distribution for the P-CV model. The solid line in the return distribution represents the median and the dashed lines are the 25th and 75th percentiles in each month. The bottom panel shows the differences in optimal portfolio weights between the P-CV and NP-CV models. The weight differences represent the effect of including return predictability in the model.

bottom panel shows the difference between the optimal portfolio weight under the P-CV model and the optimal weight under the NP-CV model. As such, this weight difference isolates the effect of incorporating information from the DP predictor variable.

The results in Figure 3 illustrate that even statistically weak evidence of return predictability can have a large effect on the optimal portfolio weight of a Bayesian investor. The optimal weight in stocks for the NP-CV model is 46% and constant across periods. An investor who believes in the P-CV model with the DP predictor varies her weight in stocks between 3% (September 2000) and 100% (June 1932) during the sample period. The time-series standard deviation of 0.22% for $\mu_{i,\tau}$ from Table III translates into a standard deviation of 16.65% for the weight in stocks. Despite the weak evidence of predictability, the location of the predictive return distribution shifts over time from the perspective of the P-CV investor as she optimally considers information from the predictor rather than discarding it based on a statistical test.

To quantify the economic significance of the DP predictor for the P-CV investor, we calculate the difference between the investor's CERs under the optimal weights for the P-CV and NP-CV models. The annualized CER gain is 0.24%, which represents the influence on the predictive return distribution from considering the DP variable. This finding is consistent with Kandel and Stambaugh's (1996) conclusion that weak statistical evidence of return predictability can be economically important.

5.2 Single-Period Horizon Results

Table IV reports the full set of CER gains across predictors. Panel A shows results for the Goyal–Welch predictors and Panel B contains corresponding results for the new predictors. For the CV cases, we report CER gains for the P-CV investor who compares weights from the P-CV and NP-CV models. We also report the time-series mean and standard deviation of the weight differences across the P-CV and NP-CV models. The SV results make the analogous comparisons between the P-SV and NP-SV models from the perspective of the P-SV investor. The CER gains are reported in percentage per year and the weight difference statistics are reported in percentage.

Beginning with the CV CER gains, the results indicate substantial variation in economic significance across predictors. The CER gains for the Goyal–Welch predictors in Panel A of Table IV range from 0.03% (DE) to 0.68% (NTIS) per year. The mean weight differences are less than 1% in magnitude for each predictor, such that little of the economic difference between the P-CV and NP-CV models is attributable to a tendency to be more or less aggressive under one framework. Across predictors, there is a close correspondence between the CER difference and the standard deviation of the optimal weight differences under the P-CV and NP-CV models. This pattern is in line with our discussion in Section 4.2.

Many of the new predictors in Panel B of Table IV have larger CER gains compared with those in Panel A. The smallest CER gain is 0.55% (DOW) and the largest is 4.36% (VRP) per year. Among the twelve new predictors, CER gains that exceed 1.00% per year are achieved by seven predictors: PLS, GAP, NOS, TAIL, GP, VRP, and LJV. The large CER gains for the new predictors reflect substantial variation in optimal weights, as evidenced by time-series standard deviations of weight differences that range from 30.73% to 90.21%. These findings confirm that the new predictors have considerable economic value to investors who believe in the CV framework studied by Kandel and Stambaugh (1996).

Comparisons of these initial CER gains with those from the SV cases in Table IV demonstrate that the economic significance of a given predictor can differ substantially depending on the investor's views on volatility. Focusing on extreme examples, the CER gain for TBL is 0.34% for the P-CV investor and 1.17% for the P-SV investor, whereas the CER gain for the PLS variable is 3.36% for P-CV versus only 0.55% for P-SV. The CER comparisons reflect differences in investor behavior, as the standard deviation of weight differences is much higher under CV for PLS and much higher under SV for TBL. These cases illustrate that considering SV can either increase or decrease the economic gains from a given predictor variable. The CER gain is larger under SV for seven of the fourteen Goyal–Welch predictors and four of the twelve new predictors.

To provide additional insights into the relative economic significance of return predictors across the volatility frameworks, Table V builds upon the analysis in Section 4.2. Three primary economic mechanisms are at work: (i) a statistical effect, (ii) an average volatility effect, and (iii) a portfolio effect. First, the volatility framework impacts how the investor learns about the predictability relation, which produces a statistical effect through the posterior of the predictive regression beta. Second, the investor's aggressiveness in betting on the conditional return forecast in a typical period depends on her views of how risky the stock market is on average, which produces an average volatility effect. Third, a

Table IV. In-sample CER gains and weight differences associated with return predictability

The table reports CER gains and weight difference statistics that reflect the economic significance of return predictability in models with CV or SV. The CV cases compare the P-CV model with the NP-CV model for each predictor and the SV cases compare the P-SV and NP-SV models. For each specification, we report the CER difference (i.e., $\Delta CER^{is}_{i,j}$) and the time-series mean (i.e., $\overline{\omega^*_{i,\tau} - \omega^*_{j,\tau}}$) and the time-series standard deviation [i.e., $\sigma(\omega^*_{i,\tau} - \omega^*_{j,\tau})$] of the difference in conditional portfolio weights. The CER gains are expressed in percent per year and the weight difference statistics are reported as percentages.

	CV			SV		
Predictor	$\Delta CER^{is}_{i,j}$	$\overline{\omega^*_{i,\tau} - \omega^*_{j,\tau}}$	$\sigma(\omega^*_{i,\tau} - \omega^*_{j,\tau})$	$\Delta CER^{is}_{i,j}$	$\overline{\omega^*_{i,\tau} - \omega^*_{j,\tau}}$	$\sigma(\omega^*_{i,\tau} - \omega^*_{j,\tau})$
Panel A: Goyal and Welch (2008) predictors						
DP	0.24	−0.34	16.65	0.39	−0.34	34.31
DY	0.53	−0.26	24.70	0.35	−0.71	32.20
EP	0.41	−0.36	21.89	0.26	−0.83	26.90
DE	0.03	−0.75	5.66	0.02	0.24	7.02
SVAR	0.20	−0.66	14.97	0.23	−1.44	13.67
BM	0.51	0.02	24.34	0.23	1.65	24.99
NTIS	0.68	−0.73	27.94	0.53	−0.21	41.52
TBL	0.34	−0.31	19.82	1.17	0.78	52.08
LTY	0.17	0.17	14.00	0.83	1.70	45.87
LTR	0.32	−0.74	19.13	0.83	−0.29	47.54
TMS	0.26	−0.71	17.22	0.19	−1.03	22.86
DFY	0.05	−0.78	7.28	0.06	−0.31	11.56
DFR	0.27	−0.74	17.69	0.01	−0.15	4.21
INFL	0.11	−0.77	11.03	0.60	−0.88	35.84
Panel B: New predictors						
PLS	3.36	1.30	62.07	0.55	−2.69	30.53
EG	0.81	−0.60	37.84	1.44	2.17	56.07
GAP	2.36	0.25	66.09	2.11	2.45	79.25
NOS	1.28	−0.46	47.24	1.52	1.24	72.46
DOW	0.55	−0.85	30.73	0.53	−2.11	38.06
TAIL	1.03	−0.54	41.89	0.52	−2.50	35.58
COR	0.72	−0.63	35.08	0.89	−1.38	55.21
SII	0.81	−1.20	35.68	0.72	−0.42	57.05
GP	1.70	−0.97	53.58	1.24	−0.65	75.04
OIL	0.65	−1.18	33.16	1.13	0.66	55.59
VRP	4.36	2.70	90.21	3.45	−0.79	78.02
LJV	1.65	−0.90	52.19	0.87	1.30	35.76

portfolio effect reflects the time-series relation between a predictor and market volatility. If a given predictor tends to make its most extreme predictions during highly volatile times, information in the predictor is less valuable to an investor.

Table V reports statistics that reflect the three effects described above. For each volatility model, we report the square root of the average squared weight difference from

Table V. Attribution of portfolio weight differences

The table reports statistics relating to the attribution of weight differences to mean components and residual components. The CV cases compare the P-CV model with the NP-CV model for each predictor and the SV cases compare the P-SV and NP-SV models. For each specification, we report the mean component $\left(\frac{\overline{\sigma^{-2}}}{\gamma}\sqrt{\frac{1}{T}\sum_{\tau=1}^{T}(\mu_{i,\tau}-\mu_{j,\tau})^2}\right)$, which estimates the impact on portfolio weights of time variation in expected return, and the total variation of the moment-based weight approximations $\left(\sqrt{\frac{1}{T}\sum_{\tau=1}^{T}\left(\frac{\mu_{i,\tau}}{\gamma\sigma_{i,\tau}^2}-\frac{\mu_{j,\tau}}{\gamma\sigma_{j,\tau}^2}\right)^2}\right)$. We calculate the ratio of total variation to mean component. The table also shows the absolute values of ratios of the predictive betas (β) and the precisions ($\overline{\sigma^{-2}}$) across the P-SV and P-CV models. The mean component and total variation statistics are reported in percentages.

	CV			SV			Absolute ratios (|SV/CV|)	
Predictor	Mean Component	Total variation	Ratio	Mean component	Total variation	Ratio	Predictive betas	Precisions
Panel A: Goyal and Welch (2008) predictors								
DP	16.57	16.52	1.00	30.47	36.29	1.19	0.95	1.90
DY	24.55	24.49	1.00	27.37	33.91	1.24	0.58	1.90
EP	21.87	21.74	0.99	26.15	28.12	1.08	0.63	1.89
DE	5.50	5.50	1.00	7.09	6.53	0.92	0.65	1.91
SVAR	15.27	14.84	0.97	49.68	13.91	0.28	1.65	1.91
BM	24.36	24.15	0.99	26.99	26.57	0.98	0.56	1.90
NTIS	28.06	27.75	0.99	43.88	47.14	1.07	0.81	1.92
TBL	19.78	19.73	1.00	57.94	56.31	0.97	1.51	1.94
LTY	13.86	13.84	1.00	45.90	50.04	1.09	1.70	1.94
LTR	19.17	18.96	0.99	46.74	48.94	1.05	1.27	1.92
TMS	17.09	17.03	1.00	21.93	23.49	1.07	0.67	1.90
DFY	7.35	7.18	0.98	19.25	11.44	0.59	1.29	1.91
DFR	17.71	17.51	0.99	3.84	3.49	0.91	0.11	1.91
INFL	11.06	10.92	0.99	45.97	36.88	0.80	2.17	1.91
Panel B: New predictors								
PLS	67.50	61.77	0.92	73.99	32.35	0.44	0.57	1.84
EG	38.22	37.43	0.98	66.91	57.64	0.86	1.25	1.40
GAP	65.74	65.31	0.99	77.67	83.01	1.07	0.84	1.40
NOS	47.36	46.84	0.99	65.89	77.41	1.17	0.94	1.48
DOW	30.75	30.38	0.99	44.29	38.83	0.88	0.95	1.46
TAIL	41.77	41.61	1.00	40.88	36.11	0.88	0.65	1.48
COR	35.12	34.69	0.99	50.53	56.52	1.12	0.94	1.49
SII	35.55	35.37	1.00	35.95	69.12	1.92	0.66	1.52
GP	53.39	53.16	1.00	48.32	89.88	1.86	0.61	1.47
OIL	33.34	32.79	0.98	65.41	59.62	0.91	1.23	1.59
VRP	107.11	86.69	0.81	172.06	85.45	0.50	1.01	1.57
LJV	55.56	50.99	0.92	79.61	36.98	0.46	0.82	1.67

Equation (20) and the square root of the mean component from Equation (21). The square root of the average squared weight difference in Table V closely corresponds to the standard deviation of optimal weights in Table IV. We also report the ratio of total variation to mean component for each predictor. A ratio greater than (less than) one indicates that the actual variation in portfolio weights is greater than (less than) what would be expected based solely on variation in $\mu_{i,\tau}$. That is, the portfolio effect is positive (negative) when the ratio is greater than (less than) one. Finally, Table V reports the absolute ratios of predictive betas (i.e., the posterior means of β) and precisions ($\overline{\sigma^{-2}}$) from the SV and CV models. The predictive beta ratio shows the strength of the statistical effect and the precision ratio corresponds to the average volatility effect. Both ratios help to determine the relative magnitudes of the mean components across the volatility frameworks.

We begin with the specific case of DP in Panel A of Table V. The mean component is 16.57% in the CV case, such that we would expect the time-series standard deviation of weight differences to be 16.57% based on information in $\mu_{i,\tau}$. This figure is quite similar to the total variation of 16.52%. Thus, the residual component is small and the ratio is close to one. These findings indicate that time variation in $\mu_{i,\tau}$ is the primary driver of economic value in the CV framework. Under the SV framework, the total variation of 36.29% is larger than the mean component of 30.47% and the ratio of these two figures is 1.19. Thus, the total variation in weights is larger for DP than would be expected based on variation in $\mu_{i,\tau}$ alone, such that the interaction between $\mu_{i,\tau}$ and $\sigma^2_{i,\tau}$ increases the economic significance of DP under the P-SV model.

We can also compare results across the P-CV and P-SV models using the ratios of predictive betas and precisions. The DP predictive beta ratio of 0.95 indicates that expected returns vary slightly less for the P-SV model than for the P-CV model. The precision ratio of 1.90 shows that the SV investor believes that the average period is less volatile compared with the CV investor, such that the P-SV investor can usually be more aggressive in betting on return forecasts. Multiplying the CV mean component by both ratios closely approximates the SV mean component (i.e., $16.57\% \times 0.95 \times 1.90 = 29.91\% \approx 30.47\%$). The DP results indicate that the relative magnitudes of CER gains of 0.24% under P-CV and 0.39% under P-SV can be reconciled by noting that (i) the predictive betas are similar across the two models (weak statistical effect), (ii) the SV investor believes she can be more aggressive in trading on DP in the typical month because of the precision term (positive average volatility effect), and (iii) the interaction between the time series of $\mu_{i,\tau}$ and $\sigma^2_{i,\tau}$ under P-SV is beneficial to the investor (positive portfolio effect).

Analyzing the remaining predictors in Panel A of Table V reveals several notable patterns. The average precision is always much higher under the SV model, such that the average volatility effect increases the economic significance of each predictor. This effect helps to generate larger mean components for thirteen of the fourteen predictors compared with the CV case. There are some cases in which the statistical effect of estimating the predictive regression beta under alternative volatility models is important. For example, the predictive beta ratio of 0.11 for DFR is influential in producing a mean component of only 3.84% for the P-SV investor compared with 17.71% for the P-CV investor, whereas the INFL predictive beta ratio of 2.17 assists in producing a mean component of 45.97% under SV versus only 11.06% under CV. Finally, the portfolio effect from interactions between $\mu_{i,\tau}$ and $\sigma^2_{i,\tau}$, reflected by the ratio of total variation to mean component, is important in several cases. As previously noted, the economic value of DP increases in the P-SV framework due to this interaction and six of the other thirteen variables also gain. Notably, the ratio is only 0.28

for the SVAR variable, such that the actual variation in portfolio weights is much less than would be expected based on the mean component. This result is intuitive given that the SVAR predictor measures stock market variance. As such, its most extreme predictions for returns correspond to high-volatility periods in which the investor is unwilling to bet aggressively.

Panel B of Table V shows results for the new predictors. The ratio of precisions remains above one for each predictor and these positive average volatility effects assist in producing higher mean components under SV for ten of the twelve predictors. As noted in Section 4, the statistical effect of estimating the predictive regression under different volatility models is important for some predictors including PLS and GP. The most important mechanism for many of the predictors, however, is the portfolio effect observable from the ratio of total weight variation to mean component. Notably, this ratio is 0.50 or less for three of the four strongest predictors from Table II: PLS, VRP, and LJV. These low ratios provide evidence that negative portfolio effects from interactions between $\mu_{i,\tau}$ and $\sigma^2_{i,\tau}$ for these variables substantially reduce their economic significance. The SII and GP variables, on the other hand, both have ratios of total variation to mean component near two. These positive portfolio effects help to offset the negative statistical effects reflected by predictive regression beta ratios of 0.66 and 0.61 for SII and GP, respectively.

Overall, Table V provides evidence on the mechanisms that determine the CER gain for a given predictor. First, the statistical effect that predictive regression beta estimates differ across volatility models is important for several predictors. Second, the average volatility effect is always positive because the average return precision is higher under the SV framework for each predictor, such that the investor believes that she can usually bet more aggressively under this framework. Third, the portfolio effect from interactions between $\mu_{i,\tau}$ and $\sigma^2_{i,\tau}$ is important for inferences about economic significance for many of the predictors.

To illustrate the mechanisms underlying the results in Table V, we more closely examine the PLS predictor variable. Table IV shows that the P-CV investor has a CER gain of 3.36% for this variable, whereas the P-SV investor's CER gain is only 0.55%. A portion of the decline in economic value of PLS is attributable to a shift in the investor's views about the statistical evidence of return predictability that is apparent in Figure 2. Nonetheless, the statistical evidence that PLS positively forecasts market returns remains strong in the P-SV model and over 98% of posterior draws of β are positive.

Table V indicates that the more pronounced effect of SV on the economic value of PLS occurs through the portfolio effect as evidenced by the ratio of weight variation to mean component of 0.44. Figure 4 shows the relation between PLS and the conditional variance of market returns. The top panel plots the time series of the predictor variable and the bottom panel shows the time series of the annualized standard deviation of market returns implied by the posterior mean of the SV process. Consistent with past empirical work, market volatility is highly time varying with large spikes that generally correspond to times of economic uncertainty. Volatility peaked during the Great Depression, with the annualized standard deviation reaching as high as 55%. This period of extreme market volatility corresponds closely to the most extreme values of PLS. In particular, 90% of PLS observations fall within the range of –1.30 to –0.45, but the variable drops as low as –3.44 during the market volatility spike in the early 1930s.

Figure 5 shows predictive return distributions and weight differences for the P-CV and P-SV investors who consider PLS. The top panels show the median (solid line) and 25th and

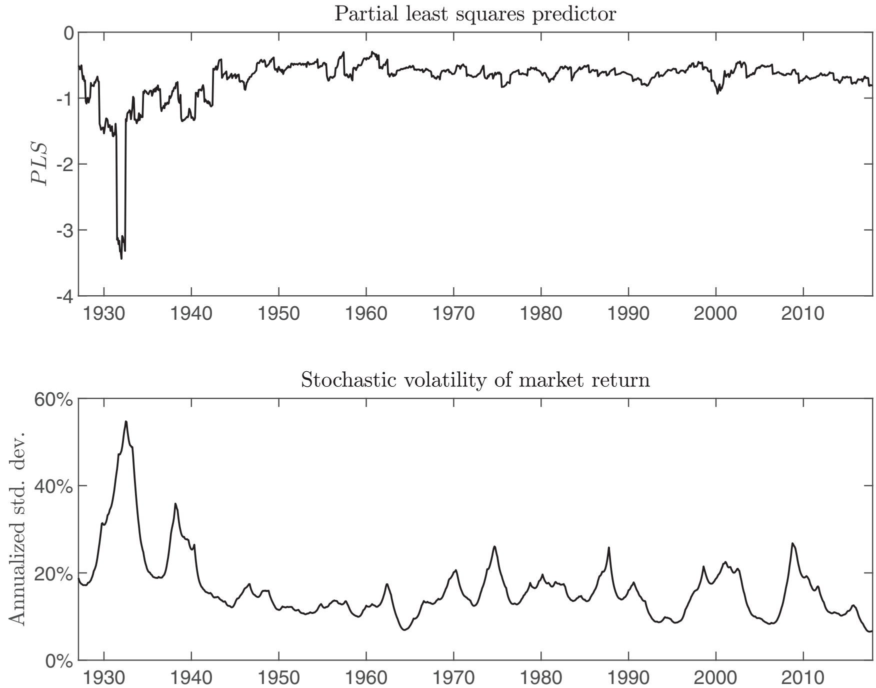

Figure 4. PLS predictor variable and SV process. The figure shows the PLS predictor variable in the top panel and the posterior mean of the annualized standard deviation of the stock market return from the SV process in the bottom panel. The monthly standard deviation is annualized by multiplying by $\sqrt{12}$.

75th percentiles (dashed lines) of the predictive return distribution from a given model. The bottom panel on the left (right) plots the difference between the optimal portfolio weight under the P-CV (P-SV) model and the optimal weight for the NP-CV (NP-SV) model. These differences represent the effect of information from PLS on the optimal portfolio weight.

The results in Figure 5 demonstrate stark differences in the predictive return distribution and optimal portfolio weights across the CV and SV cases. The most extreme differences occur during the high-volatility period in the early 1930s. In the month with the lowest conditional return forecast (January 1932), the P-CV investor aggressively bets against stocks with a weight of –392%, which reflects an expected excess return of –6.31% and a predictive standard deviation of 5.57% for the subsequent month. The P-SV investor adopts a relatively modest portfolio weight of –25% in the same month given an expected excess return of –3.19% and a high predictive standard deviation of 13.32%. The tendency for the most extreme return forecasts for PLS to line up with periods of high volatility thus reduces the P-SV investor's ability to capitalize on information in the predictor. This time-series interaction between $\mu_{i,\tau}$ and $\sigma^2_{i,\tau}$ is reflected in Table V by the low ratio of weight variation to mean component of 0.44 in the SV case.

5.3 Multi-Period Horizon Results

We now consider the economic significance of stock return predictors from the perspective of Bayesian investors with multi-period horizons. Our broad sample of predictors produces

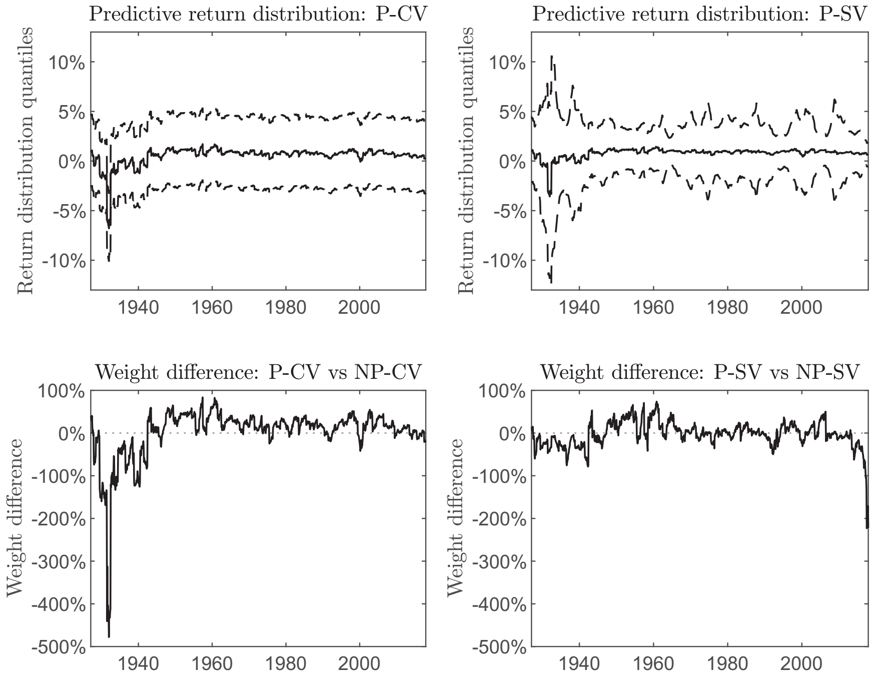

Figure 5. Predictive return distribution quantiles and effects of predictability on portfolio weight for the PLS predictor. The figure shows quantiles of the predictive return distributions for the P-CV and P-SV models in the top panels and the differences in optimal portfolio weights between the P-CV and NP-CV models and the P-SV and NP-SV models in the bottom panels in which PLS is the market return predictor variable. The solid line in the return distribution represents the median and the dashed lines are the 25th and 75th percentiles in each month. The weight differences represent the effect of including return predictability in the model for the CV and SV cases.

a wide variety of time-series properties as described in Section 3. Overall, we find the most dramatic multi-period effects for predictors with low persistence or pronounced SV in the predictor (which is often reflected by extreme skewness and kurtosis in Table I). We use the VRP predictor as an example to illustrate these economic channels in this section and we include a full analysis and discussion of all predictors in Online Appendix E. The VRP example is interesting because of the large effects on perceived economic value even with small increases in horizon, but we also find that the persistence and SV effects are important for many of the predictors.

The CER gains for VRP decline dramatically when the horizon increases from 1 month to as short as 3 months. Whereas the 1-month CER gain in the CV framework is 4.36% per year for VRP, the annualized 3-month CER gain is only 1.00%. Similarly, in the SV framework, the 1-month CER gain is 3.45% versus only 1.32% with a 3-month horizon. The VRP variable exhibits low persistence, substantial SV in the predictor process, and a tendency for extreme values within the sample period. Each of these features works against the value of the variable for longer-term Bayesian investors.

Figure 6 shows the VRP predictor variable along with the SV processes for returns and the state variable. The VRP variable displays a noticeable downward spike in October 2008, but we note that this observation post-dates the sample periods of the Bollerslev,

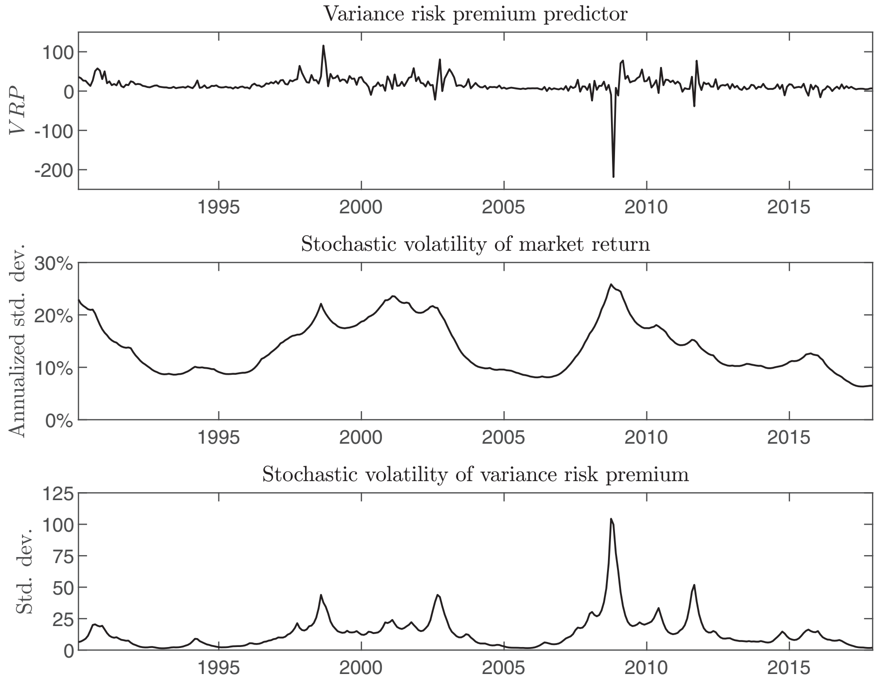

Figure 6. VRP predictor variable and SV processes. The figure shows the VRP predictor variable in the top panel, the posterior mean of the annualized standard deviation of the stock market return from the SV process in the middle panel, and the posterior mean of the standard deviation of the VRP predictor variable from the SV process in the bottom panel. The monthly standard deviation of market returns is annualized by multiplying by $\sqrt{12}$.

Tauchen, and Zhou (2009) and Drechsler and Yaron (2011) studies such that the initial evidence for VRP in the literature is not driven by the outlier. Nonetheless, this observation contributes to a pattern observed throughout the sample that the conditional volatility of VRP is highly variable. There is also a tendency for spikes in the volatility of VRP to coincide with spikes in market volatility.

Figure 7 plots predictive return distributions and weight differences for 1-month and 3-month horizons. Panel A (Panel B) shows results for the CV (SV) models. The predictive return distribution plots report the median and 25th and 75th percentiles of the distribution of 1-month or 3-month cumulative returns. The weight differences are the difference between optimal weights for the P-CV and NP-CV models in Panel A and the difference between optimal weights for the P-SV and NP-SV models in Panel B.

The 1-month, CV results show that the investor aggressively shifts her portfolio weights in response to changes in VRP. The weight in stocks for the P-CV investor ranges from –656% (November 2008) to 512% (September 1998) compared with the NP-CV weight of 76%. The optimal weights for the 3-month P-CV investor are muted relative to the 1-month results. The weight in stocks varies between –297% and 285% during the sample period (the NP-CV optimal weight is 75%). The most important impact of the longer horizon for this investor is the lack of persistence in VRP as indicated by the low posterior mean of 0.28 for β_x, such that any variation in expected return is perceived to be short-

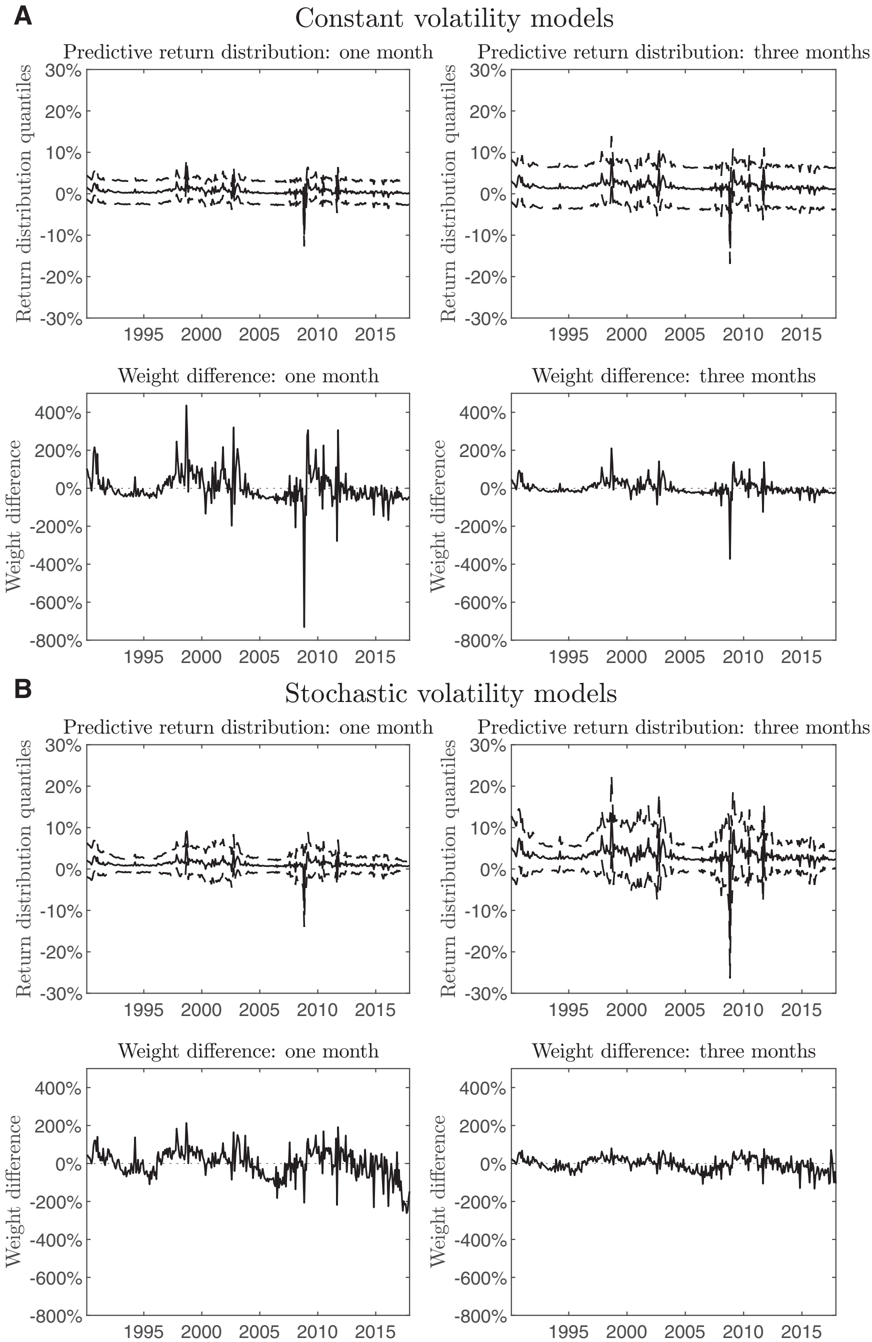

Figure 7. Predictive return distribution quantiles and effects of predictability on portfolio weight for the VRP predictor. The figure shows quantiles of the predictive return distributions and weight differences for the CV models in Panel A and the SV models in Panel B in which VRP is the market return predictor variable. In Panel A (Panel B), the top panels show quantiles of the predictive return

lived and primarily affects only the first month of the holding period. The more moderate positions taken by the 3-month P-CV investor are reflected by the relatively low CER gain of 1.00% compared with 4.36% for the 1-month investor.

The results in Panel B of Figure 7 for the SV cases also show large horizon effects. The weight difference for the 3-month P-SV investor is noticeably devoid of large spikes, although there is variation in the optimal weight that produces a CER gain of 1.32%. The mechanism at work in the SV case is different from that in the CV case. After considering SV, the investor perceives VRP to be more persistent compared with the CV model given the posterior mean of 0.63 for β_x, such that the variation in the conditional 3-month expected return suggests that utility gains could be large. However, the optimal portfolio weights are still relatively moderate for the 3-month P-SV investor because of effects of horizon on beliefs about risk. In particular, investing for multi-period horizons is risky from the perspective of the Bayesian investor when the predictor variable takes extreme values and is highly volatile. Intuitively, even if the conditional expected return for the first month in the holding period is high, the volatile predictor implies that the expected returns in months 2 and 3 are highly uncertain. Thus, a positive bet on the high conditional expected return runs the risk that the expected return moves against the bet during the holding period. In the most extreme periods for VRP, the 3-month predictive return distributions are marked by high variance and kurtosis as a result of this additional uncertainty.[14] In the months in which the predictor makes its most extreme forecasts of the conditional expected return, the investor moderates her positions and avoids making large bets given the high perceived risk. This behavior is reflected by the relatively low 3-month CER gain.

6. Out-of-Sample Results

The results in Sections 4 and 5 correspond to an in-sample design. The advantage of this approach is that we consider information from the full sample in characterizing how the relations between predictors and volatility impact economic significance (see, e.g., Inoue and Kilian, 2005). Following the influential study of Goyal and Welch (2008), the literature on aggregate stock market return predictability now places considerable emphasis on out-of-sample tests. Such tests offer a complementary characterization of the value of return

Figure 7. Continued
distribution for the P-CV (P-SV) model at the 1-month and 3-month horizons. The solid line in the return distribution represents the median and the dashed lines are the 25th and 75th percentiles in each month. The bottom panels of Panel A (Panel B) show the differences in optimal portfolio weights between the P-CV and NP-CV (P-SV and NP-SV) models at the 1-month and 3-month horizons. The weight differences represent the effect of including return predictability in the model for each case.

14 Pástor and Stambaugh (2012); Avramov, Cederburg, and *Lučivjanská* (2018); and Carvalho, Lopes, and McCulloch (2018) demonstrate that uncertainty about future expected returns has a substantial positive effect on predictive variance over long horizons. We complement the findings of these studies by showing large effects over short horizons for predictors with low persistence and high volatility.

predictability by considering the performance of hypothetical investors who rely on predictors to make real-time forecasts and portfolio allocations.

In this section, we present results from out-of-sample tests. As detailed in Section 2.4.2, these tests employ an expanding-window design with an initial training period of 240 months. For each combination of model and predictor, we estimate the model for each month of the out-of-sample period and determine the optimal portfolio weight using data that would have been available to a real-time investor. Given the full time series of out-of-sample portfolio weights for each model and predictor, we construct a time series of realized portfolio returns and compute the CER for this return series under power utility. Our analysis focuses on the CER differences for the P-CV and NP-CV models and for the P-SV and NP-SV models. We assess statistical significance of the CER differences using a bootstrap approach as detailed in Online Appendix A.

Prior research suggests that out-of-sample tests of aggregate stock market return predictability impose a high bar for success. Goyal and Welch (2008) consider a comprehensive set of predictors and find that almost all exhibit poor performance in forecasting mean returns out of sample and very few offer utility benefits for real-time investors. Several papers published subsequent to Goyal and Welch (2008) present evidence that individual predictor variables perform well in out-of-sample exercises, but the empirical methods differ across studies. Moreover, Goyal, Welch, and Zafirov (2021) examine many of these recently proposed predictors using an extended sample and a unified empirical framework and find that most variables offer no forecasting or investment benefits in real time.

There is some reason for ex ante optimism in our setting, as the BVAR framework accounts for realistic features of the real-time portfolio choice problem, including parameter uncertainty and SV in market returns and predictors. Johannes, Korteweg, and Polson (2014) employ a similar econometric framework and show economically significant portfolio benefits for real-time investors who rely on the payout yield. They emphasize the importance of accounting for time-varying volatility and estimation risk in achieving these out-of-sample gains. Our results offer an important contribution by incorporating these features in tests for a much broader set of predictors. We also expect the economic effects related to predictor properties highlighted in Sections 4 and 5 to play an important role in the out-of-sample analysis. That is, predictors exhibiting fat tails and taking extreme values in high volatility periods are likely to be of more limited economic value to investors.

For each predictor, Table VI reports the out-of-sample CER gains at the 1-month investment horizon. The CV results are CER differences in percentage per year for investors using the P-CV model relative to those using the NP-CV model. We also report the time-series standard deviation of the weight differences across the P-CV and NP-CV models. The SV results provide an analogous comparison for the P-SV and NP-SV models. Online Appendix F contains supplementary empirical results for the out-of-sample analysis, including distributional statistics for the out-of-sample returns for each model and predictor, out-of-sample Sharpe ratio gains, and out-of-sample CER gains for investors with longer holding periods. Our design choice to start the out-of-sample period 240 months after the start of data availability for each predictor is consistent with prior studies (see, e.g., Goyal and Welch, 2008; Goyal, Welch, and Zafirov, 2021), but produces a short out-of-sample evaluation window for some of the new predictors (e.g., VRP and LJV). We demonstrate in Online Appendix F, however, that the conclusions in Table VI are robust to using a shorter initial training sample of 120 months.

Table VI. Out-of-sample CER gains

The table reports out-of-sample CER gains ($\Delta\text{CER}^{os}_{i,j}$) that reflect the economic significance of return predictability in models with CV or SV. The CV cases compare the P-CV model with the NP-CV model for each predictor and the SV cases compare the P-SV and NP-SV models. The CER gains are expressed in percentage per year. We assess statistical significance of the CER gains using a bootstrap approach. The *P*-value corresponds to the one-sided test of the null hypothesis that the gain is less than zero. For each specification, we also report the time-series standard deviation (in percentage) of the conditional weight difference [i.e., $\sigma(\omega^*_{i,\tau} - \omega^*_{j,\tau})$]. The holding period for the portfolio strategies is 1 month.

		CV			SV		
Predictor	OS start	$\Delta\text{CER}^{os}_{i,j}$	*P*-value	$\sigma(\omega^*_{i,\tau} - \omega^*_{j,\tau})$	$\Delta\text{CER}^{os}_{i,j}$	*P*-value	$\sigma(\omega^*_{i,\tau} - \omega^*_{j,\tau})$
Panel A: Goyal and Welch (2008) predictors							
DP	1947:02	–0.44	0.819	17.92	2.17	0.036	48.65
DY	1947:02	–0.86	0.809	42.39	3.06	0.035	66.43
EP	1947:02	–0.89	0.812	37.60	2.07	0.024	41.15
DE	1947:02	0.13	0.346	11.46	0.36	0.259	24.18
SVAR	1947:02	0.16	0.153	5.47	1.23	0.000	12.45
BM	1947:02	–1.12	0.940	32.84	1.01	0.149	35.28
NTIS	1947:02	–0.10	0.538	23.81	–0.28	0.615	40.56
TBL	1947:02	–0.13	0.586	24.34	3.86	0.011	64.21
LTY	1947:02	–0.60	0.800	25.40	3.07	0.033	56.46
LTR	1947:02	0.00	0.499	21.16	1.91	0.077	64.69
TMS	1947:02	0.29	0.257	19.06	0.84	0.301	44.70
DFY	1947:02	–0.19	0.966	4.91	0.66	0.047	17.22
DFR	1947:02	–0.07	0.573	14.01	–0.07	0.535	18.73
INFL	1947:02	–0.09	0.655	8.66	1.44	0.014	30.76
Panel B: New predictors							
PLS	1947:02	1.02	0.067	23.75	0.24	0.346	31.81
EG	1959:06	0.03	0.473	24.32	0.80	0.094	24.24
GAP	1968:01	2.78	0.126	76.93	2.50	0.091	70.19
NOS	1978:03	–0.23	0.572	48.90	0.54	0.320	59.46
DOW	1980:02	–1.11	0.848	35.70	–0.35	0.653	37.17
TAIL	1983:02	0.77	0.250	37.86	0.88	0.140	34.78
COR	1983:04	–2.56	0.877	44.73	–3.57	0.977	53.83
SII	1993:02	1.27	0.034	24.87	0.23	0.345	29.68
GP	1995:02	2.17	0.208	72.78	1.60	0.180	61.35
OIL	2003:05	–12.91	0.983	62.47	–3.65	0.816	95.56
VRP	2010:02	3.18	0.181	59.92	3.22	0.192	64.45
LJV	2016:07	–4.30	1.000	7.76	–4.00	1.000	33.72

Most of the Goyal–Welch predictors in Panel A of Table VI lead to CER losses under the CV framework. Ten of the fourteen CER differences are negative, with an extreme value of –1.12% per year for BM. Of the four positive CER gains under CV, none is statistically significant at conventional levels and the largest value is just 0.29% per year (TMS).

These results are consistent with those of Goyal and Welch (2008) who find similarly weak out-of-sample performance for these predictors. Their frequentist asset allocation exercise is similar to our CV comparison, but does not account for parameter uncertainty.

Relative to the Goyal–Welch predictors, the new predictors in Panel B of Table VI generate more extreme investment results under the CV framework. The CER gains range from –12.91% for OIL to 3.18% for VRP. The new predictors tend to be somewhat more successful than the Goyal–Welch predictors are, as seven of the twelve CER gains are positive. Just two of the CER differences (PLS and SII), however, are statistically significant at the 10% level.

The CV results in Table VI reveal another important distinction between in-sample and out-of-sample inferences. Whereas the discussion in Section 4.2 implies a direct connection between weight variation and CER gains based on the predictive return distribution, the out-of-sample tests condition on subsequent return realizations. Aggressive bets carry the potential for extreme performance realizations in this setting. Table VI indicates that larger weight variation tends to be associated with CER differences that are large in absolute magnitude, whether they be gains (e.g., the 2.78% CER gain for GAP, which has a 76.93% standard deviation of weight differences) or losses (e.g., the 12.91% CER loss for OIL with a 62.47% standard deviation).

The most striking result in Table VI is that accounting for SV in returns and predictors leads to substantially better out-of-sample investment performance. In line with the results and discussion in Section 5.2, the SV investors tend to trade more aggressively compared with the CV investors for the Goyal–Welch predictors as evidenced by the standard deviations of weight differences, and these bets tend to pay off. The P-SV model generates a larger CER than does the NP-SV model for twelve of the fourteen Goyal–Welch predictors. Across these fourteen predictors, the CER gains are as high as 3.86% per year for TBL, and the worst performing model (NTIS) results in a CER loss of just 0.28% per year. Nine of the CER gains for the Goyal–Welch predictors are statistically significant at the 10% level. These results are surprising given the notoriously poor out-of-sample performance for these predictors shown in prior work. They also highlight the importance of incorporating realistic features into the portfolio choice problem, most notably SV and parameter uncertainty.

In Panel B, the new predictors fare reasonably well under the SV framework, but the performance is considerably less impressive relative to that of the Goyal–Welch predictors. Eight of the twelve CER gains in Panel B are positive. There are, however, three variables that produce annualized CER differences below –3.00% (COR, OIL, and LJV). Among the new predictors, only EG and GAP generate out-of-sample CER gains that are statistically significant at the 10% level.[15] Nonetheless, the totality of the results in Table VI generalizes

15 We acknowledge that some predictors could be economically insignificant on a stand-alone basis, but still contribute to economic value in a multi-predictor setting. To provide a preliminary examination of this question, we estimate "kitchen sink" versions of the P-CV and P-SV models using the set of fifteen predictors that have data spanning the full sample period. We focus on out-of-sample performance in this setting given the tendency of highly parameterized models to overfit the data in sample and Martin and Nagel's (2021) emphasis on the importance of out-of-sample tests in high-dimensional prediction problems. We find that the P-CV model produces an annualized CER gain of –5.30% per year at the one-month horizon relative to the NP-CV model. This result is consistent with the poor out-of-sample performance of kitchen sink models in prior work (see, e.g., Goyal and Welch, 2008). Perhaps surprisingly, the P-SV model leads to an out-of-sample CER gain of 3.20%

Johannes, Korteweg, and Polson's (2014) finding of the importance of accounting for time-varying volatility to a broad set of return predictors.

7. Summary

Table VII provides a summary of our analysis. The first three columns of the table indicate whether or not each predictor exhibits empirical properties that we have argued are signals of potential economic value: (i) low kurtosis, (ii) high persistence, and (iii) a tendency to take on moderate values during times of high stock market volatility. The remaining columns summarize the statistical performance of the predictors in OLS regressions (Table II) and the economic performance of the predictors based on in-sample CER gains (Table IV and Table E.IV in the Online Appendix) and out-of-sample CER gains (Table VI and Table F.V in the Online Appendix). Table VII reveals four key takeaways.

First, in the classic 1-month, CV setting, there is a strong in-sample relation between statistical and economic performance. The OLS R^2 from a predictive return regression is a strong indicator of economic value, and even weak statistical evidence can produce non-trivial utility gains.

Second, even with a 1-month horizon, accounting for SV in returns can drive a wedge between the statistical and economic performance of a given predictor. As we formalize in the paper, SV impacts the economic significance of many predictors through a statistical effect, an average volatility effect, and a portfolio effect. The portfolio effect is most pronounced for predictors exhibiting high kurtosis and a tendency to take extreme values in periods of high stock market volatility. Several predictors exhibit these properties (e.g., BM, PLS, VRP, and LJV) and the economic value of these variables under SV (in-sample SV1 column in Table VII) is muted relative to that under CV (in-sample CV1 column).

Third, in multi-period settings, the persistence level of predictor variables produces strong effects on expected return and risk. For example, the estimated half-lives of the AR processes followed by the NOS, DOW, TAIL, OIL, and VRP predictors are less than 1 year (i.e., $\hat{\beta}_x < 0.944$ in Table II) and these predictors see large drops in economic value going from the in-sample CV1 case to the in-sample CV3 case in Table VII.

Fourth, the importance of predictor properties on economic outcomes extends to out-of-sample settings, particularly when the investor accounts for SV. We confirm the well-established result that the predictors from Goyal and Welch (2008) lead to poor out-of-sample performance under the CV framework. Many of these variables, however, produce large CER gains under SV. The real-time performance of the new predictors in the SV setting is noticeably less consistent, as the distributional properties of these variables can limit their economic value to investors.

If we define an economically successful predictor as one that delivers CER gains in excess of 0.5% per year in both in-sample and out-of-sample tests at multiple horizons while accounting for SV, then six of the twenty-six predictors meet this threshold: TBL, LTY, EG, GAP, NOS, and GP. Notably absent from this list are several new predictors with impressive statistical performance.

relative to the NP-SV model. We are cautious to draw broad conclusions about the generalizability of the success of SV models in multiple-predictor settings. Based on these results, however, the importance of accounting for SV and parameter uncertainty on equity premium prediction with multiple predictors represents an interesting question for future work.

Table VII. Summary of results

The table provides a summary of properties for stock market return predictor variables, the statistical performance of these variables in standard predictive regressions, and the economic performance of these variables based on in-sample and out-of-sample asset allocation exercises. The columns under the "predictor properties" heading detail whether or not a given predictor follows a distribution with fat tails, has low persistence, or takes extreme values in high-volatility periods. The column under the "OLS" heading details whether or not a given predictor generates an R^2 value above 0.5% in an OLS predictive regression. The columns under the "in-sample results" and "out-of-sample results" headings summarize in-sample and out-of-sample CER gains that reflect the economic significance of return predictability in models with CV or SV at horizons of 1 or 3 months. The CV cases compare the P-CV model with the NP-CV model for each predictor and the SV cases compare the P-SV and NP-SV models.

Predictor properties	OLS	In-sample results	Out-of-sample results
✓:Condition satisfied ✗:Condition violated P1: $\text{Kurt}(x_t) < 5.0$ P2: $\beta_x > 0.944$ P3: $\rho(\lvert x_t - \bar{x}_t\rvert, r_t^2) < 0.2$	✓ : $R^2 > 0.5\%$ ✗ : $R^2 \leq 0.5\%$	✓ : $\Delta\text{CER}^{\text{is}}_{i,j} > 0.5\%$ +(−): $\Delta\text{CER}^{\text{is}}_{i,j} > 0.5\%$ and 0.5% larger (smaller) than CV1 case	✓ : $\Delta\text{CER}^{\text{os}}_{i,j} > 0.5\%$ ✗ : $\Delta\text{CER}^{\text{os}}_{i,j} < 0.0\%$

Predictor	P1	P2	P3	OLS	CV1	CV3	SV1	SV3	CV1	CV3	SV1	SV3
Panel A: Goyal and Welch (2008) predictors												
DP	✓	✓	✓	✗	—	—	—	—	✗	✗	✓	✓
DY	✓	✓	✓	✗	✓	✓	—	—	✗	✗	✓	✓
EP	✗	✓	✓	✗	—	—	—	—	✗	✗	✓	✓
DE	✗	✓	✗	✗	—	—	—	—	—	—	—	—
SVAR	✗	✗	✗	✗	—	—	—	—	—	—	✓	✓
BM	✓	✓	✗	✗	✓	—	—	—	✗	✗	✓	—
NTIS	✗	✓	✓	✓	✓	✓	✓	—	✗	✗	✗	✗
TBL	✓	✓	✓	✗	—	—	✓$^{+}$	✓$^{+}$	✗	✗	✓	✓
LTY	✓	✓	✓	✗	—	—	✓$^{+}$	✓$^{+}$	✗	✗	✓	✓
LTR	✗	✗	✓	✗	—	—	✓$^{+}$	—	—	—	✓	✗
TMS	✓	✓	✓	✗	—	—	—	—	—	—	✓	✓
DFY	✗	✓	✗	✗	—	—	—	—	✗	✗	✓	✗
DFR	✗	✗	✗	✗	—	—	—	—	✗	—	✗	✗
INFL	✗	✗	✓	✗	—	—	✓	—	✗	—	✓	—
Panel B: New predictors												
PLS	✗	✓	✗	✓	✓	✓$^{-}$	✓$^{-}$	✓$^{-}$	✓	✓	—	✓
EG	✗	✗	✓	✓	✓	✓	✓$^{+}$	✓	—	✗	✓	✓
GAP	✓	✓	✓	✓	✓	✓	✓	✓$^{-}$	✓	✓	✓	✓
NOS	✓	✗	✓	✓	✓	✓$^{-}$	✓	✓$^{-}$	✗	✓	✓	✓
DOW	✓	✗	✗	✗	✓	—	✓	—	✗	✗	✗	✗
TAIL	✓	✗	✓	✓	✓	✓$^{-}$	✓$^{-}$	—	✓	—	✓	—
COR	✓	✗	✗	✓	✓	✓	✓	✓	✗	✗	✗	✗
SII	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	—	—
GP	✓	✓	✓	✓	✓	✓	✓	✓$^{-}$	✓	✓	✓	✓
OIL	✗	✗	✓	✓	✓	—	✓	—	✗	✗	✗	✗
VRP	✗	✗	✗	✓	✓	✓$^{-}$	✓$^{-}$	✓$^{-}$	✓	✓	✓	✗
LJV	✗	✓	✗	✓	✓	✓	✓$^{-}$	✓$^{-}$	✗	✗	✗	✗

8. Conclusion

We evaluate the economic significance of stock market return predictors from the perspective of Bayesian investors while accounting for realistic features of the data and investors' capital allocation decisions. Our work complements Kandel and Stambaugh's (1996) result that even weak statistical evidence of predictability is economically important by showing that there are several predictors for which strong statistical findings correspond to limited economic significance.

Supplementary Material

Supplementary data are available at *Review of Finance* online.

References

Avramov, D. (2002): Stock return predictability and model uncertainty, *Journal of Financial Economics* 64, 423–458.

Avramov, D. (2004): Stock return predictability and asset pricing models, *Review of Financial Studies* 17, 699–738.

Avramov, D. and Wermers, R. (2006): Investing in mutual funds when returns are predictable, *Journal of Financial Economics* 81, 339–377.

Avramov, D. and Zhou, G. (2010): Bayesian portfolio analysis, *Annual Review of Financial Economics* 2, 25–47.

Avramov, D., Kosowski, R., Naik, N. Y., and Teo, M. (2011): Hedge funds, managerial skill, and macroeconomic variables, *Journal of Financial Economics* 99, 672–692.

Avramov, D., Barras, L., and Kosowski, R. (2013): Hedge fund return predictability under the magnifying glass, *Journal of Financial and Quantitative Analysis* 48, 1057–1083.

Avramov, D., Cederburg, S., and Lučivjanská, K. (2018): Are stocks riskier over the long run? Taking cues from economic theory, *Review of Financial Studies* 31, 556–594.

Banegas, A., Gillen, B., Timmermann, A., and Wermers, R. (2013): The cross section of conditional mutual fund performance in European stock markets, *Journal of Financial Economics* 108, 699–726.

Barberis, N. (2000): Investing for the long run when returns are predictable, *Journal of Finance* 55, 225–264.

Bollerslev, T., Engle, R. F., and Nelson. D. B. (1994): ARCH models, in R. F. Engle, and D. L. McFadden (eds.), *Handbook of Econometrics*, Vol. 4, North-Holland, pp. 2959–3038.

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

Bollerslev, T., Todorov, V., and Xu, L. (2015): Tail risk premia and return predictability, *Journal of Financial Economics* 118, 113–164.

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

Campbell, J. Y. (1991): A variance decomposition for stock returns, *The Economic Journal* 101, 157–179.

Carvalho, C. M., Lopes, H. F., and McCulloch, R. E. (2018): On the long-run volatility of stocks, *Journal of the American Statistical Association* 113, 1050–1069.

Chen, L. and Zhang, L. (2011): Do time-varying risk premiums explain labor market performance?, *Journal of Financial Economics* 99, 385–399.

Christopherson, J. A., Ferson, W. E., and Glassman, D. A. (1998): Conditioning manager alphas on economic information: another look at the persistence of performance, *Review of Financial Studies* 11, 111–142.

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

Cochrane, J. H. (2011): Discount rates, *Journal of Finance* 66, 1047–1108.

Cooper, I. and Priestley, R. (2009): Time-varying risk premiums and the output gap, *Review of Financial Studies* 22, 2801–2833.

Cremers, K. J. M. (2002): Stock return predictability: a Bayesian model selection perspective, *Review of Financial Studies* 15, 1223–1249.

Negro, M. D. and Primiceri, G. E. (2015): Time varying structural vector autoregressions and monetary policy: a corrigendum, *Review of Economic Studies* 82, 1342–1345.

Drechsler, I. and Yaron, A. (2011): What's vol got to do with it?, *Review of Financial Studies* 24, 1–45.

Driesprong, G., Jacobsen, B., and Maat, B. (2008): Striking oil: another puzzle?, *Journal of Financial Economics* 89, 307–327.

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

Fleming, J., Kirby, C., and Ostdiek, B. (2001): The economic value of volatility timing, *Journal of Finance* 56, 329–352.

Fleming, J., Kirby, C., and Ostdiek, B. (2003): The economic value of volatility timing using "realized" volatility, *Journal of Financial Economics* 67, 473–509.

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

Goyal, A., Welch, I., and Zafirov, A. (2021): A comprehensive look at the empirical performance of equity premium prediction II. Working paper, University of Lausanne.

Huang, D. and Kilic, M. (2019): Gold, platinum, and expected stock returns, *Journal of Financial Economics* 132, 50–75.

Inoue, A. and Kilian, L. (2005): In-sample or out-of-sample tests of predictability: which one should we use?, *Econometric Reviews* 23, 371–402.

Johannes, M., Korteweg, A., and Polson, N. (2014): Sequential learning, predictability, and optimal portfolio returns, *Journal of Finance* 69, 611–644.

Johnson, T. L. (2019): A fresh look at return predictability using a more efficient estimator, *Review of Asset Pricing Studies* 9, 1–46.

Jones, C. S. and Tuzel, S. (2013): New orders and asset prices, *Review of Financial Studies* 26, 115–157.

Kandel, S. and Stambaugh, R. F. (1996): On the predictability of stock returns: an asset allocation perspective, *Journal of Finance* 51, 385–424.

Kelly, B. and Jiang, H. (2014): Tail risk and asset prices, *Review of Financial Studies* 27, 2841–2871.

Kelly, B. and Pruitt, S. (2013): Market expectations in the cross-section of present values, *Journal of Finance* 68, 1721–1756.

Klein, R. W. and Bawa, V. S. (1976): The effect of estimation risk on optimal portfolio choice, *Journal of Financial Economics* 3, 215–231.

Li, J. and Yu, J. (2012): Investor attention, psychological anchors, and stock return predictability, *Journal of Financial Economics* 104, 401–419.

Martin, I. and Nagel, S. (2021): Market efficiency in the age of big data, Working paper, London School of Economics and Political Science.

Pástor, L. and Stambaugh, R. F. (2000): Comparing asset pricing models: an investment perspective, *Journal of Financial Economics* 56, 335–381.

Pástor, L., and Stambaugh, R. F. (2012): Are stocks really less volatile in the long run?, *Journal of Finance* 67, 431–477.

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

Pollet, J. M. and Wilson, M. (2010): Average correlation and stock market returns, *Journal of Financial Economics* 96, 364–380.

Primiceri, G. E. (2005): Time varying structural vector autoregressions and monetary policy, *Review of Economic Studies* 72, 821–852.

Rapach, D. E. and Zhou, G. (2022): Asset pricing: Time-series predictability. *Oxford Research Encyclopedia of Economics and Finance*, forthcoming.

Rapach, D. E., Ringgenberg, M. C., and Zhou, G. (2016): Short interest and aggregate stock returns, *Journal of Financial Economics* 121, 46–65.

Rossi, P. E., Allenby, G. M., and McCulloch, R. (2005): *Bayesian Statistics and Marketing*, John Wiley & Sons, West Sussex, UK.

Shanken, J. and Tamayo, A. (2012): Payout yield, risk, and mispricing: a Bayesian analysis, *Journal of Financial Economics* 105, 131–152.

Stambaugh, R. F. (1999): Predictive regressions, *Journal of Financial Economics* 54, 375–421.

Tu, J. and Zhou, G. (2010): Incorporating economic objectives into Bayesian priors: portfolio choice under parameter uncertainty, *Review of Financial Studies* 45, 959–986.

Wachter, J. A. (2010): Asset allocation, *Annual Review of Financial Economics* 2, 175–206.

Wachter, J. A. and Warusawitharana, M. (2009): Predictable returns and asset allocation: should a skeptical investor time the market?, *Journal of Econometrics* 148, 162–178.

Wachter, J. A. and Warusawitharana, M. (2015): What is the chance that the equity premium varies over time? Evidence from regressions on the dividend-price ratio, *Journal of Econometrics* 186, 74–93.