The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXIX, NO. 2 • APRIL 2014

Sequential Learning, Predictability, and Optimal Portfolio Returns

MICHAEL JOHANNES, ARTHUR KORTEWEG, and NICHOLAS POLSON*

ABSTRACT

This paper finds statistically and economically significant out-of-sample portfolio benefits for an investor who uses models of return predictability when forming optimal portfolios. Investors must account for estimation risk, and incorporate an ensemble of important features, including time-varying volatility, and time-varying expected returns driven by payout yield measures that include share repurchase and issuance. Prior research documents a lack of benefits to return predictability, and our results suggest that this is largely due to omitting time-varying volatility and estimation risk. We also document the sequential process of investors learning about parameters, state variables, and models as new data arrive.

EQUITY RETURN PREDICTABILITY is widely considered a stylized fact: theory indicates that expected returns should time-vary and numerous studies find supporting evidence. For example, Lettau and Ludvigson (2001, p. 842) argue "it is now widely accepted that excess returns are predictable by variables such as dividend-price ratios, earnings-price ratios, dividend-earnings ratios, and an assortment of other financial indicators." Evidence for predictable volatility is so strong that it is rarely debated, with predictability introduced via short-run persistence and long-run mean reversion. This predictability should be important for investors when making portfolio decisions, as investors should "time" the investment set, increasing allocations when expected returns are high and/or volatility is low.

A surprising recent finding questions the evidence for equity return predictability, and suggests that there are no out-of-sample benefits to investors from exploiting this predictability when making optimal portfolio decisions.

*Johannes is at the Graduate School of Business, Columbia University. Korteweg is at the Graduate School of Business, Stanford University. Polson is at the Booth School of Business, University of Chicago. We thank Cam Harvey (the Editor); the Associate Editor; an anonymous referee; Martijn Cremers; Darrell Duffie; Wayne Ferson; Pulak Ghosh; Stefan Nagel; and seminar participants at Columbia, Duke, Rice, the University of North Carolina, USC Marshall, Yale School of Management, the University of Chicago, the 2012 Society of Quantitative Analysts Meeting, the 2009 AFA meetings, the 2008 Conference on Modeling and Forecasting Economic and Financial Time Series with State Space models at Sveriges Riksbank, the 2009 CREATES conference in Skagen, Denmark, the 2009 SOFiE conference in Lausanne, the 2009 CIREQ-CIRANO Financial Econometrics Conference in Montreal, and the 2009 Quantitative Methods in Finance Symposium at UT Austin for helpful comments. We thank Ravi Pillai for excellent computing support. All errors are our own.

DOI: 10.1111/jofi.12121

Goyal and Welch (2008, p. 1456) find that "the evidence suggests that most models are unstable or even spurious. Most models are no longer significant even in-sample.... Our evidence suggests that the models would not have helped such an investor" seeking to use the predictability when forming portfolios. Intuitively, the conclusion is that, while there may be some evidence for predictability, it is so weak as to be of no practical use for investors.

This paper revisits this issue and finds new results reconciling these seemingly contradictory findings. We find strong evidence that investors can use predictability to improve out-of-sample portfolio performance provided they incorporate a number of sensible features into their optimal portfolio problems. In addition to incorporating expected return predictability, investors must account for time-varying volatility and estimation risk when forming portfolios. Our results are not inconsistent with Goyal and Welch (2008), as we also find no benefits to expected return predictability using the standard approach, which assumes constant volatility regression models and investors who ignore estimation risk.

Intuitively, an *ensemble* of additional features is needed because each feature provides only a marginal performance improvement. For example, time-varying volatility is important as it is both highly variable and predictable, and estimation risk is important because there is substantial uncertainty over the nature of the predictability (Brennan (1998), Stambaugh (1999), Barberis (2000)). Ignoring either of these components provides a misleading view of the risks. Incorporating estimation risk or time-varying volatility alone does not, however, generate statistically significant out-of-sample improvements for the standard predictability model. One way to interpret our results is that careful modeling requires accounting for all of the first-order important features, such as predictable expected returns, time-varying volatility, and parameter uncertainty. Thus, there is no single "silver bullet" generating out-of-sample gains.

Our empirical experiment is straightforward. We consider a Bayesian investor who (1) uses models incorporating yield-based expected return predictors and stochastic volatility, (2) learns about the models, parameters, and state variables sequentially in real time, revising beliefs via Bayes's rule as new data arrive, and (3) computes predictive return distributions and maximizes expected utility accounting for all sources of uncertainty. We focus on a single predictor variable, the dividend yield, but consider two measures ofthis variable: the traditional cash dividends measure and a version incorporating equity repurchases and issuance. The latter measure was introduced by Boudoukh et al. (2007), who find that net payout yield is a stronger predictor of equity returns. Overall, we confront our investor with the same learning problems faced by econometricians, a problem suggested by Hansen (2007).

To implement the Bayesian portfolio problem, we need to characterize the posterior distribution at each point in time throughout our sample. We use particle filters to tackle this difficult sequential learning problem. Particle filters are a recursive Monte Carlo approach that generate approximate samples from the posterior distribution that can then be used to generate draws from the predictive return distributions to compute optimal portfolio holdings. Particle filters are the dominant approach for sequential state or parameter inference across a range of fields.

After solving the learning problem, our investor maximizes expected constant relative risk aversion (CRRA) utility over terminal wealth for different time horizons, from one month to two years. Ideally, one would solve the recursive long-horizon portfolio problem with intermediate learning, but this is infeasible with multiple unknown parameters.[1] Given these portfolios, we compute out-of-sample portfolio returns, summarizing performance using standard metrics such as Sharpe ratios and certainty equivalent returns (CEs). CEs are a more relevant benchmark than Sharpe ratios given power utility. This procedure generates a time series of realized *fully out-of-sample* returns for various models and data sets (cash dividend yields and net payout yields). To evaluate the statistical significance, we simulate returns assuming a given model, for example, constant means and variances, and evaluate models with various forms of predictability to see if the Sharpe ratios or CEs are statistically different from those generated by simpler model specifications.[2]

Empirically, our first set of results indicates that none of the constant volatility models generates statistically significant out-of-sample improvements compared to a model with constant means and variances. This finding implies that accounting for parameter uncertainty and using the Boudoukh et al. (2007) net payout yield predictor does not provide statistically significant benefits, assuming constant volatility. This result is consistent with Goyal and Welch (2008), but goes one step further and implies that just accounting for parameter uncertainty (i.e., being Bayesian) does not generate statistically significant improvements. In some cases, timing based on expected return predictability using the traditional cash dividend measure performs worse than using a model with constant means and variances (accounting for parameter uncertainty in both cases). This result is robust for all risk aversion cases and across all investment horizons considered.

Our main result is that incorporating an ensemble of factors significantly improves out-of-sample performance. A specification with predictable expected returns generated by net payout yields and stochastic volatility, when used by an investor who accounts for estimation risk, generates statistically significant (at the 5% level) improvements in CEs and Sharpe ratios. This holds for all risk aversions and investment horizons, where significance is measured either against a model with constant means and variances or against a model with constant means and time-varying volatilities.

[1] The Bellman equation generated by the fully dynamic problem is high-dimensional. Essentially, each unknown state and parameter has sufficient statistics, and thus the dimensionality of the Bellman equation is roughly equal to twice the number of unknown parameters and states, on the order of 25 dimensions for even the simplest models. Solving this problem is not feasible with current computing capabilities.

[2] Although our investor is Bayesian, there are no methodological problems in evaluating the out-of-sample returns generated by a Bayesian investor using classical statistical techniques. We thank the Associate Editor and a referee for suggesting this experiment.

The effects are economically large. For example, in a model with constant means and variances, a Bayesian investor with a risk aversion of four generates an annualized CE yield of 4.77% and a monthly Sharpe ratio of 0.089 (annualized Sharpe ratio of about 0.31). In the general model using net payout yield as the predictor and incorporating stochastic volatility, the investor generates a CE yield of 6.85% and a Sharpe ratio of 0.155 (annualized, 0.54). The 2% difference in CE yields generates extremely large gains when compounded over a sample of almost 80 years. The Sharpe ratios are more than 70% higher. The results are even stronger for long-horizon investors. Together, the results indicate that an ensemble of factors generates statistically and economically significant improvements.

Models with constant expected returns and time-varying volatility do not generate statistically significant returns, even if the investor accounts for estimation risk. Thus, we find no evidence for a pure volatility timing effect. To our knowledge, there is no published evidence for volatility timing based on aggregate equity returns over long sample periods. Fleming, Kirby, and Ostdiek (1998) consider a multivariate asset problem using data from 1982 to 1996 and study time-varying second moments, which include correlations. They also use significant in-sample information about average returns. Yan (2005) considers a problem with many individual stocks and factor stochastic volatility. Bandi, Russell, and Zhu (2008) consider multiple individual stocks and volatility timing using intraday high-frequency equity returns.

If the cash dividend yield is used instead of net payout yields along with time-varying volatility, we find statistically strong improvements, but not as large as the improvements generated by the net payout yield measure. Thus, the net payout measure provides additional portfolio benefits, as it is a stronger predictor of equity returns. We also consider a drifting coefficients specification, but this model generally performs on par or slightly worse than models with constant predictability.

We also summarize the investor's real-time learning about parameters, states, and models. We find evidence that learning can take a significant amount of time, which should not be surprising given the persistence of volatility, dividend yields, and expected returns. This does explain, however, why incorporating estimation risk can be important, as there is significant uncertainty over parameter estimates even after observing decades of data. We also discuss the model learning problem, quantifying how an investor learns about the relative merits of competing models.

We connect our approach to the recent results in Pástor and Stambaugh (2012) on term structures of predictive variances. They find that predictive return volatility does not necessarily fall as the time horizon increases, in contrast to what would happen with i.i.d. returns and in contrast to popular belief. They document this feature in the context of a "predictive system" in which the relationship between the predictor variables and expected returns is imperfect. Predictive volatilities can increase with horizon due to parameter and state variable uncertainty. We perform the same experiments as Pástor and Stambaugh (2012) and find similar results in our models. Although our models are not a formal imperfect predictive system, our results indicate that the increasing predictive volatility as a function of time horizon is a more general feature, as it appears in models other than those considered in Pástor and Stambaugh (2012).

The rest of the paper is as follows. Section I describes the standard approach for evaluating predictability via out-of-sample returns, the models we consider, and our methodology. Section II reports our results on sequential inference, including parameter estimates and model probabilities, and the out-of-sample portfolio results. Section III concludes.

I. Evaluating Predictability via Out-of-Sample Portfolio Performance

A. The Standard Approach

The standard approach considers a model of the form

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

where r_{t+1} are monthly log excess returns on the CRSP value-weighted portfolio, x_t is a predictor variable, ε_t^r is a mean-zero constant variance error term, and the coefficients α, β, and σ^2 are fixed but unknown parameters. The dividend yield is the most commonly used predictor, defined as the natural logarithm of the previous year's cash payouts divided by the current price. Standard full-sample statistical tests for predictability estimate the models on a long historical sample commonly starting in 1927.[3] It is possible to incorporate multiple predictors, but this paper follows the bulk of the literature and focuses on univariate models.

Although statistical significance is important for testing theories, measures of economic performance, such as the performance of optimal portfolios out-of-sample, are arguably more appropriate and require that investors could identify and take advantage of predictability in real time. Standard implementations of out-of-sample portfolio experiments such as Goyal and Welch (2008) use regression models like the one above combined with the assumption of normally distributed errors to form optimal portfolios. An investor finds portfolio weights between aggregate equities and the risk-free rate by maximizing one-period expected utility, assuming a power or CRRA utility function, and using the predictive distribution of returns induced by the regression model. Initial parameter estimates are estimated using a training sample, and are reestimated as new data arrive. Point estimates for the parameters are used to predict future returns, which is called the plug-in method. As mentioned

[3] For recent results in this area and extensive citations, see Shiller (1981), Hodrick (1992), Stambaugh (1999), Xia (2001), Avramov (2002), Cremers (2002), Ferson, Sarkissian, and Simin (2003), Lewellen (2004), Torous, Valkanov, and Yan (2004), Campbell and Yogo (2006), Cochrane (2008), Ang and Bekaert (2007), Campbell and Thompson (2008), Lettau and van Nieuwerburgh (2008), Pástor and Stambaugh (2009), and Shanken and Tamayo (2012). Shanken and Tamayo (2012) allows volatility to depend deterministically on the predictor variable and is in that sense closest to our paper, but it does not allow for stochastic volatility.

earlier, Goyal and Welch (2008) find no benefits to an investor who follows this procedure using a wide range of predictors. In particular, they find no benefits for the "classic" predictor variable, cash dividend yield. Wachter and Warusawitharana (2009) consider a Bayesian multiasset portfolio problem with long-term bonds, aggregate equity returns, and the risk-free rate. They find out-of-sample benefits for a highly informative prior, but no benefits for other priors. They provide no evidence that the gains are due to timing expected returns in stocks, and their optimal portfolios maintain large short positions in long-term bonds, which implies that they have a large negative bond risk premium. Thus, any gains are partially due to bond and not stock positions. The gains they find are quite modest, relative to the gains we document below. Wachter and Warusawitharana (2012) consider a related problem with dividend yield timing, but do not consider out-of-sample returns.

Prima facie, there are multiple reasons to suspect that the standard approach might perform poorly out-of-sample. First, the regression model above ignores important first-order features of equity returns. Most notably, the constant volatility assumption is in strong contrast to observed data, since equity return volatility time-varies. Ignoring this variation could cause optimal portfolios based solely on time-varying expected returns to perform poorly. Moreover, power utility specifications are sensitive to fat tails in the return distribution, a feature absent in the constant volatility, normally distributed shock regression specification, but present in models with time-varying volatility.

Second, the standard approach ignores the fact that the parameters determining the equity premium, α and β, are estimated with a significant amount of error. Indeed, the debate about predictability has received so much attention in part because the predictability evidence, while compelling, is still quite weak. By ignoring estimation risk or parameter uncertainty, the standard implementation understates the total uncertainty, as perceived by an investor. Kandel and Stambaugh (1996) and Barberis (2000) document the important role of parameter uncertainty when forming optimal portfolios.

Third, the linear regression model assumes that the relationship between x_t and r_{t+1} is time-invariant. Theoretically, certain asset pricing models, such as Menzly, Santos, and Veronesi (2004) or Santos and Veronesi (2006), imply that the relationship between the equity premium and x_t time-varies. Empirically, Paye and Timmerman (2006), Lettau and van Nieuwerburgh (2008), Henkel, Martin, and Nardari (2011), and Dangl and Halling (2012) find evidence for time variation in the relationship between returns and common predictors.

Fourth, most out-of-sample implementations based on expected return predictability focus on the dividend yield, which measures payouts via cash dividends. As argued by Boudoukh et al. (2007), an expanded measure of payout including share repurchases is a far more effective predictor. In fact, they argue that there is no evidence that cash dividends are a significant predictor but net payout is strongly significant. For all of these reasons, it may not at all be surprising that the standard approach performs poorly out-of-sample.

The goal of this paper is to introduce extensions to deal with these features and reevaluate the out-of-sample performance. The next section introduces the models and our empirical approach.

B. Our Approach

B.1. Models

We consider a number of extensions to the baseline regression model. The first allows volatility to vary over time

$$r_{t+1} = \alpha + \beta x_t + \sqrt{V^r_{t+1}}\varepsilon^r_{t+1}, \tag{2}$$

where V^r_{t+1} evolves via a log-volatility specification (Jacquier, Polson, and Rossi (1994, 2005))

$$\log\left(V^r_{t+1}\right) = \alpha_r + \beta_r \log\left(V^r_t\right) + \sigma_r \eta^r_{t+1}. \tag{3}$$

In choosing the log-specification, the goal is to have a parsimonious specification ensuring that volatility is stochastic, positive, and mean-reverting. Volatility predictability arises from its persistent but mean-reverting behavior.

Time-varying volatility has direct and indirect effects on optimal portfolios. The direct effect is through the time variation in the investment set generated by stochastic and mean-reverting volatility, as investors “time” volatility, increasing or decreasing equity allocations as volatility changes over time. This effect is ignored in constant volatility regression models. There is also an indirect effect because time-varying volatility implies that the signal-to-noise ratio for learning about expected return predictability varies over time. To see this, note that the time t log-likelihood function for the parameters controlling equity premium, conditional on volatility, is

$$\ln\left(L\left(r_{t+1}, x_t, V^r_{t+1}|\alpha, \beta\right)\right) = c_{t+1} - \frac{1}{2}\frac{\left(r_{t+1} - \alpha - \beta x_t\right)^2}{V^r_{t+1}}, \tag{4}$$

where c_{t+1} does not depend on the parameters. In models with constant volatility, $V^r_t = \sigma^2$, the amount of information regarding expected return predictability is constant over time. When volatility time-varies, the information content varies with V^r_t. When V^r_t is high, there is little information about expected returns, and thus the signal-to-noise ratio is low. Conversely, when V^r_t is low, the signal-to-noise ratio is high. This is, of course, the usual generalized least squares (GLS) versus ordinary least squares (OLS) problem that vanishes asymptotically, but can be important in this setting due to small-sample issues generated by the high persistence of x_t and the relatively low signal-to-noise ratio.

The stochastic volatility specification has an additional important feature for optimal portfolios: it generates fat-tailed return distributions. The distribution of returns in equation (2) is normally distributed, conditional on V^r_{t+1} and the parameters, but the marginal and predictive distributions of returns that integrate out the unobserved volatilities are a scale mixture of normals, which has fat tails. In addition to fitting the variation in volatility, time-varying volatility is a long-standing explanation for fat tails (see, for example, Rosenberg (1972)). The continuous-time literature finds that stochastic volatility alone cannot generate enough kurtosis to fit the observed return data at high frequencies, such as daily, but at lower frequencies, such as monthly, stochastic volatility models generate excess kurtosis that is consistent with the observed returns. This issue is discussed in more detail below. We assume that the volatility shocks are independent of returns.[4]

We also allow the regression coefficient on the predictor variable to vary over time. As mentioned above, some theories imply that this coefficient varies, and there is also empirical evidence suggesting that the loading on predictors such as the dividend-price ratio varies over time (Lettau and van Nieuwerburgh (2008), Henkel, Martin, and Nardari (2011), Dangl and Halling (2012)). This extension posits that β_t, the regression coefficient, is a mean-reverting process with mean β_0 and autocovariance β_β. The model is

$$r_{t+1} = \alpha + \beta_0 x_t + \beta_{t+1} x_t + \sqrt{V^r_{t+1}}\varepsilon^r_{t+1}, \tag{5}$$

$$\beta_{t+1} = \beta_\beta \beta_t + \sigma_\beta \varepsilon^\beta_{t+1}, \tag{6}$$

where ε^β_{t+1} is i.i.d. normal. It is common to assume that β_t moves slowly, consistent with values of β_β close to one and σ_β relatively small. Alternatively, a Markov switching process would allow for abrupt changes in the states. The drifting coefficient specification is related to Pástor and Stambaugh (2009), who consider latent specifications of the conditional mean, where the shocks in the conditional mean are correlated with returns and predictor variables. We discuss the connections in greater detail below.

Based on Stambaugh (1986), we model x_t as a persistent but mean-reverting process

$$x_{t+1} = \alpha_x + \beta_x x_t + \sqrt{V^x_{t+1}}\varepsilon^x_{t+1}, \tag{7}$$

where $\beta_x < 1$, $\mathrm{corr}(\varepsilon^r_t, \varepsilon^x_t) = \rho$, and V^x_{t+1} is the time-varying variance of dividend yields. We assume a standard log-specification for V^x_{t+1}, $\log(V^x_{t+1}) = \alpha_v + \beta_v \log(V^x_t) + \sigma_v \eta^v_{t+1}$, where the errors are standard normal.

[4] This significantly simplifies implementation, as the mixture approximation of Kim, Shephard, and Chib (1998) can be used in econometric implementation. The leverage effect is often motivated by negative skewness in equity returns; for example, at a daily frequency, the skewness of aggregate equity is typically about −2 (see Andersen, Benzoni, and Lund (2002)). The skewness is much less significant at monthly frequencies, roughly −0.49, and is not statistically different from zero. We estimated a specification incorporating a leverage effect using the full sample of returns, and the point estimate was only −0.11, which is much smaller (in absolute value) than typically found at the daily frequency (e.g., Eraker, Johannes, and Polson (2003) and Jacquier, Polson, and Rossi (2005) find values of around −0.5).

Incorporating a mean-reverting process for x_t is particularly important for optimal portfolios formed over long horizons, which we consider in addition to monthly horizons. As noted by Stambaugh (1999), mean reversion in x_t generates skewness in the predictive distribution of returns at longer horizons.

We consider the following specifications:

- The "CV-CM" model has constant mean (CM) and constant variance (CV). This is a benchmark model with no predictability (i.e., equation (1) with $\beta = 0$).
- The "CV" model has constant variance but time-varying expected returns. In equations (2) and (7), this is the special case with $V^r_{t+1} = \sigma^2$ and $V^x_{t+1} = \sigma^2_x$.
- The "CV-OLS" model is the same model as CV, but implemented using OLS with all data up to time t.
- The "CV-rolling OLS" model is the same model as CV, but implemented using OLS and a 10-year rolling window of data.
- The "CV-DC" model is a constant volatility model with drifting regression coefficients. In equations (5) through (7), this is the special case with $V^r_{t+1} = \sigma^2$ and $V^x_{t+1} = \sigma^2_x$.
- The "SV-CM" model is a stochastic volatility model with constant mean, that is, equation (2) with $\beta = 0$, which implies that the equity premium is constant.
- The "SV" model is a stochastic volatility model with time-varying expected returns generated by equations (2) and (7).
- The "SV-DC" model is the most general specification with stochastic volatility and predictability driven by the drifting coefficients model in equations (5) through (7).

All of the models are implemented using a Bayesian approach to account for parameter uncertainty, with the exception of the CV-rolling OLS and CV-OLS implementations, which condition on point estimates. We use these to highlight the impact of parameter uncertainty on out-of-sample performance. We focus on payout yield as a single predictor, but use two measures of yield: the traditional cash dividend yield measure and a more inclusive measure of total payouts via the net payout measure of Boudoukh et al. (2007), which includes share issuances and repurchases.

More general specifications are certainly possible, but our goal is not to find the most general econometric specification. Rather, our goal is to model a number of data features that are important for optimal portfolios including predictability in expected returns, time-varying volatility, contemporaneous correlation between dividend growth shocks and returns, and drifting coefficients. More general specifications could incorporate nonnormal return shocks, leverage effects, additional predictor variables, and a factor stochastic covariance structure for dividend growth and returns.[5] A large literature modeling

[5] We did consider an extension with correlated volatility shocks that captures the fact that the aggregate dividend growth and equity return volatility are significantly correlated. The out-of-

aggregate market volatility develops more involved continuous-time specifications with multiple volatility factors and nonnormal jump shocks. These models are typically implemented using daily or even higher frequency data; it would be very difficult to identify the above features with lower frequency monthly data.

In addition, adding economic restrictions generated by present-value calculations such as those in Koijen and Van Binsbergen (2010) may also improve the model's performance.[6] These extensions add additional parameters and, more importantly, significantly complicate econometric implementation, making sequential implementation extremely difficult. It is important to note that, if our models have any gross misspecification, it should be reflected in poor out-of-sample returns.

B.2. Inference

We consider a Bayesian investor learning sequentially over time about the unobserved variables, parameters, state variables, and models. Notationally, let $\{\mathcal{M}_j\}_{j=1}^{M}$ denote the models under consideration. In each model there is a vector of unknown static parameters θ and a vector of unobserved state variables $L_t = (V_t^r, V_t^x, \beta_t)$. The observed data consist of a time series of returns and predictor variables $y^t = (y_1, \ldots, y_t)$, where $y_t = (r_t, x_t)$. The Bayesian solution to the inference problem is $p(\theta, L_t, \mathcal{M}_j | y^t)$, the posterior distribution, for each model specification at each time point. The marginal distributions $p(\theta|\mathcal{M}_j, y^t)$, $p(L_t|\mathcal{M}_j, y^t)$, and $p(y^t|\mathcal{M}_j)$ summarize parameter, state variable, and likelihood-based model inference, respectively. The marginal likelihood function, $p(y^t|\mathcal{M}_j)$, summarizes model fit and allows likelihood-based model comparisons without resorting to common approximations such as the Akaike or Bayesian Information Criterion (AIC or BIC).

Out-of-sample experiments require estimation of each model at each time period $t = 1, \ldots, T$. This real-time or sequential perspective significantly magnifies any computational difficulties associated with estimating latent variable models. For full-sample inference, Markov Chain Monte Carlo (MCMC) methods are commonly used, but they are too computationally burdensome to use sequentially. To sample from the posterior distributions, we use a Monte Carlo approach called particle filtering.

Particle filters discretize the support of the posterior, and, as shown by Johannes and Polson (2006) and Carvalho et al. (2010, 2011), work well for

sample portfolio results were similar to the other stochastic volatility specifications. We thank the referee for suggesting the specification and the exercise.

[6] Koijen and Van Binsbergen's (2010) approach introduces nonlinear parameter restrictions related to present values via a Campbell and Shiller (1988) log-linearization, assuming that underlying shocks have constant volatility. They expand around stationary means, assuming that the conditional variances are constant. This approach is difficult to implement sequentially. Parameters used in the approximations, such as stationary means, are unknown. In addition, the nonlinear parameter constraints significantly complicate Bayesian inference, as the models are no longer conjugate.

parameter and state variable inference in many models with latent states such as log-stochastic volatility models. Particle filters are fully sequential methods: after summarizing the posterior at time t, there is never any need to use past data as particle filters only use new data to update previous beliefs. Because of their sequential nature, particle filters are computationally much faster than alternatives such as repeated implementation of MCMC methods. This is the main advantage, but there is an associated cost: particle filtering methods are not as general or robust as MCMC methods. An Internet Appendix provides an overview of particle filters as well as details of our filtering algorithm, which is an extension of the methods developed in Johannes and Polson (2006) and Carvalho et al. (2010, 2011).[7]

B.3. Optimal Portfolios and Out-of-Sample Performance Measurement

When making decisions, a Bayesian investor computes expected utility using the predictive distribution, which automatically accounts for estimation risk. The posterior distribution quantifies parameter uncertainty or estimation risk. This can be contrasted with frequentist statistics, where parameters are fixed but unknown quantities and not random variables, and therefore one cannot define concepts like parameter uncertainty.

Our investor maximizes expected utility over terminal wealth T periods in the future, assuming that the wealth at the beginning of each period is \$1,

$$\max_{\{\omega\}} \mathrm{E}_t[U(W_{t+T})|\mathcal{M}_j, y^t], \tag{8}$$

where wealth evolves from t to $t+T$ via

$$W_{t+T} = W_t \cdot \prod_{\tau=1}^{T} \left[(1-\omega_{t+\tau-1})\exp(r^f_{t+\tau}) + \omega_{t+\tau-1}\exp\left(r^f_{t+\tau} + r_{t+\tau}\right)\right], \tag{9}$$

and $r^f_{t+\tau}$ is a zero-coupon default-free log bond yield for the period between time $t+\tau-1$ and $t+\tau$. The portfolio weight on equities is $\omega_{t+\tau-1}$, and is allowed to vary over the investment horizon. We consider a range of horizons T from one month ($T=1$) to two years ($T=24$).[8] In the long-horizon problems, we allow investors to rebalance their portfolios every year, as in Barberis (2000). We bound portfolio weights at -2 and $+3$, which primarily impacts the OLS models (CV-OLS and CV-rolling OLS). The out-of-sample returns from these models are much worse with uncapped weights. The portfolio weights for the other models are more stable and rarely hit the upper or lower bounds.

We consider a power utility investor

$$U(W_{t+T}) = \frac{(W_{t+T})^{1-\gamma}}{1-\gamma}. \tag{10}$$

[7] The Internet Appendix may be found in the online version of the article on the *Journal of Finance* website.

[8] Previous versions of the paper considered horizons up to 10 years, with similar results.

Expected utility is calculated for each model

$$\mathrm{E}_t[U(W_{t+T})|\mathcal{M}_j, y^t] = \int U(W_{t+T})p(W_{t+T}|\mathcal{M}_j, y^t)dW_{t+T}, \qquad (11)$$

using equation (9) and the predictive distribution of returns,

$$p(r_{t+\tau}|\mathcal{M}_j, y^t) = \int p(r_{t+\tau}|\theta, L_t, \mathcal{M}_j, y^t)p(\theta, L_t|\mathcal{M}_j, y^t)d\theta dL_t. \qquad (12)$$

Calculating expected utility in this manner, rational Bayesian investors take all of the relevant uncertainty into account by averaging across the unknown parameters and latent state variables, using the posterior distribution $p(\theta, L_t|\mathcal{M}_j, y^t)$.[9]

Marginalization alters the conditional return distribution, increasing variance and generating fat tails. To see this, consider a stochastic volatility specification where the predictive distribution is

$$p(r_{t+1}|\mathcal{M}_j, y^t) = \int p(r_{t+1}|\theta, V_t, \mathcal{M}_j, y^t)p(V_t|\theta, \mathcal{M}_j, y^t)p(\theta|\mathcal{M}_j, y^t)d\theta dV_t, \qquad (13)$$

where $p(r_{t+1}|\theta, V_t, \mathcal{M}_j, y^t)$ is the normally distributed *conditional* return distribution, $p(V_t|\theta, \mathcal{M}_j, y^t)$ is the filtered distribution of the stochastic variance, and $p(\theta|\mathcal{M}_j, y^t)$ is the parameter posterior distribution at time t.[10] Although the return distribution is conditionally normal, the predictive distribution has higher variance and fat tails generated by marginalizing out the uncertainty in volatility and the other parameters. Thus, although the shocks are normally distributed, predictive return distributions are generally nonnormal. This nonnormality is minor in constant volatility models, but substantial when volatility time-varies. This is important for fitting fat-tailed aggregate equity returns. Our power utility specification takes into account the conditional nonnormalities, which can be important (see also Brandt et al. (2005), Harvey and Siddique (2000), and Harvey et al. (2010)).

At each time period, our investor chooses portfolio weights to maximize expected utility. The investor holds the assets for a given period, realizes gains and losses, updates posterior distributions, and then recomputes optimal portfolio weights (rebalancing annually rather than monthly for investment horizons

[9] Well-known problems can arise regarding the existence of expected utility under Bayesian learning with power utility. This is due to the t-distributed predictive distribution of returns, which implies that expected utility can be infinite since the t-distribution's moment-generating function does not exist (see, for example, Kandel and Stambaugh (1996)). This result is not a generic result, but rather is caused by the standard conjugate inverse gamma prior. For example, expected power utility is finite if the volatility parameter, σ, is bounded at any finite value with a truncated inverse gamma prior. This was recently noted in Bakshi and Skoulakis (2010) in a different setting. To allow for shorting and portfolio weights greater than one, we bound the monthly return distribution between -100% and 100%. Empirically, these bounds are never hit in our simulations.

[10] An earlier version of this paper also considered optimal portfolios generated by model averaging, taking into account the fact that there are multiple models.

beyond one year). This procedure is repeated for each time period, generating a time series of out-of-sample returns. Using this time series, standard summary statistics such as CE yields and Sharpe ratios are computed to summarize portfolio performance. For some models, we document strong disagreement between CE yields and Sharpe ratios, which is generated by the fact that Sharpe ratios do not take into account tail behavior. Given that the portfolios are formed by maximizing a power utility specification, CE yields are more appropriate.

B.4. Evaluating Statistical Significance

To assess the statistical significance of the out-of-sample return summaries, the CE yields, and Sharpe ratios, we perform extensive Monte Carlo simulations to construct finite-sample distributions of the performance statistics. Our base simulations consider a null model with no predictability—constant mean and variance—that is calibrated to match the full-sample returns. Then, given returns simulated from this null model, we estimate each of our models sequentially using the same estimation procedures that we used on the real data. We repeat this exercise 500 times for each model specification to obtain a distribution of CE yields and Sharpe ratios that we can use to assess whether the statistics obtained from the real-world data are statistically significantly higher than those generated in the null model.[11,12]

We also consider the null of a stochastic volatility model with constant means. This provides a benchmark stochastic volatility specification without time-varying expected returns, allowing us to discriminate between timing based solely on volatility and timing based jointly on expected returns and volatility. This is important because stochastic volatility, as discussed above, can have both direct and indirect effects on the optimal portfolios, the former through volatility timing and the latter via time-varying signal-to-noise ratios. As in the previous case, we simulate returns and then reestimate models for each of the 500 simulated series using the same procedures used on real data.

II. Empirical Results

We use monthly log-excess returns from the value-weighted NYSE-Amex-NASDAQ index (including distributions) minus the one-month Treasury bill rate from Ibbotson and Associates over the period 1927 to 2007:

$$r_{t+1} = \ln\left((P_{t+1} + D_{t+1})/P_t\right) - \ln(1 + r_t^f)\,. \quad (14)$$

[11] This simulation exercise is extremely computationally intensive. Estimating each of the models with latent variables (drifting coefficients or stochastic volatility models) and forming portfolios takes roughly one day on a desktop machine. We run 500 simulations for eight models for both the dividend yield and payout yield data. To perform this experiment, we used a large-scale supercomputing cluster, which, after efficiently programmed, took almost six weeks of cluster computing time.

[12] We only consider one-sided tests for improved performance. The difference between statistically equal or worse performance is not important for our purposes.

Here, D_{t+1} denotes dividends obtained during period t, and P_{t+1} is the ex-dividend price. The dividend yield regressor is constructed as the natural logarithm of the sum of the previous 12 months of dividends (from CRSP) divided by the current price, as in Cochrane (2008). The net payout measure is from Boudoukh et al. (2007), which starts in 1927 and ends in 2007. This measure includes both dividends and net equity repurchases (repurchases minus issuances) over the last 12 months, scaled by the current price, and can be obtained from the authors' website.

The choice of monthly time horizon is motivated by prior literature. Since stochastic volatility movements are often high frequency, monthly data are more informative than lower frequencies such as annual. In addition, we analyze optimal portfolio allocation problems that have typically been analyzed using data at the monthly frequency; see Kandel and Stambaugh (1996), Stambaugh (1999), or Barberis (2000). Figure 1 provides time series of the regressors, OLS regression estimates, and t-statistics. The top panel indicates that net payouts are consistently higher than cash dividends over the sample period but the two are broadly similar. Repurchases used to be quite rare but have increased since the 1980s. Overall, the net payout variable is less persistent than the cash dividend yield because firms deliberately smooth cash dividends (Brav et al. (2005)), while the net payout variable contains two additional sources of variation through issuances and repurchases.

The middle and bottom panels of Figure 1 provide OLS regression coefficient estimates and t-statistics for the null hypothesis of $\mathcal{H}_0 : \beta = 0$, sequentially through the sample. The regression estimates and t-statistics are cumulative up to time t, adding new data points as they become available (and keeping all old data points). The regression coefficients and the associated t-statistics are consistently higher for net payout yield than for cash dividends over the sample period. One source of the increased significance is the higher frequency movements in net payouts. The t-statistics change significantly over time, falling significantly in the late 1990s and increasing back to prior levels by about 2003. This pattern is consistent with the findings in Boudoukh et al. (2007).

Our Bayesian investor uses standard conjugate priors described in the Internet Appendix, which are calibrated as follows. First, we train the priors from 1927 to 1929 by regressing excess market returns on a constant and the predictor. This procedure can be viewed as assuming noninformative priors, and then updating using the likelihood function using the training sample, which results in a proper conjugate prior distribution. For the SV parameters, we run AR(1) regressions using the logarithm of squared residuals on lagged log-squared residuals. The initial volatility states are drawn from the distribution of the regression volatility estimate over the training period. For time-varying coefficient models, the return and payout ratio regressions are insufficient to pin down the priors, so we place some structure on the parameters governing the evolution of β_t. The prior on β_β is calibrated to have mean 0.9, with standard deviation 0.1, implying a high autocorrelation in β_t. The conditional means and variances are equal for all models for the first out-of-sample date. This training sample approach is commonly used to generate "objective" priors.

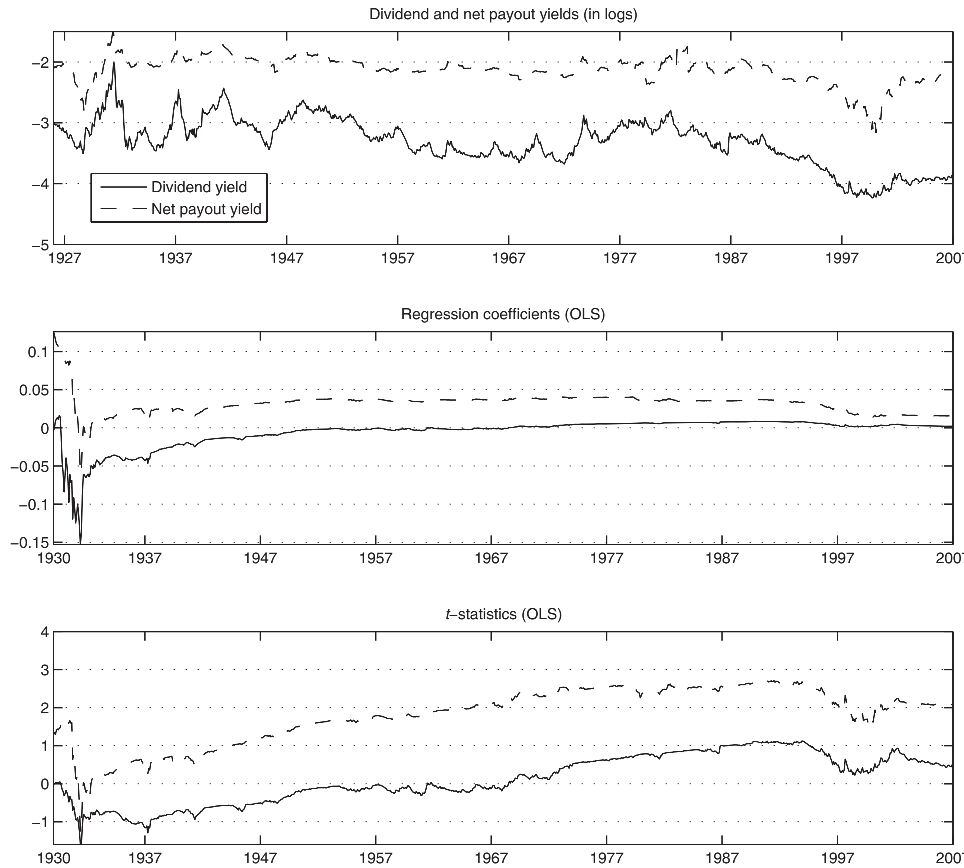

Figure 1. Sequential OLS parameter estimates. The top panel plots the time series of the two predictors, the dividend yield, and the net payout yield. The middle panel graphs OLS regression coefficients, β, of the univariate predictability regression

$$r_t = \alpha + \beta x_{t-1} + \sigma \varepsilon_t,$$

where r_t is the excess market return, the predictor x_t is either the dividend yield or net payout yield, and ε_t is distributed $\mathcal{N}(0, 1)$. We use the entire time series of excess returns, r_t, up to time t to estimate β. The bottom panel shows the t-statistics, $t(\beta)$. We use the Amihud and Hurvich (2004) method to adjust for small-sample bias.

A. Sequential Parameter Estimates and Predictive Returns

A.1. Sequential Parameter Estimates

Our approach generates parameter posteriors for each time period, for each model specification, and for both predictors. This section discusses the CV model estimated using the net payout yield measure. Results for the other models and data sets are given in the Internet Appendix. Figure 2 displays sequential summaries of the posterior distribution, reporting for each parameter

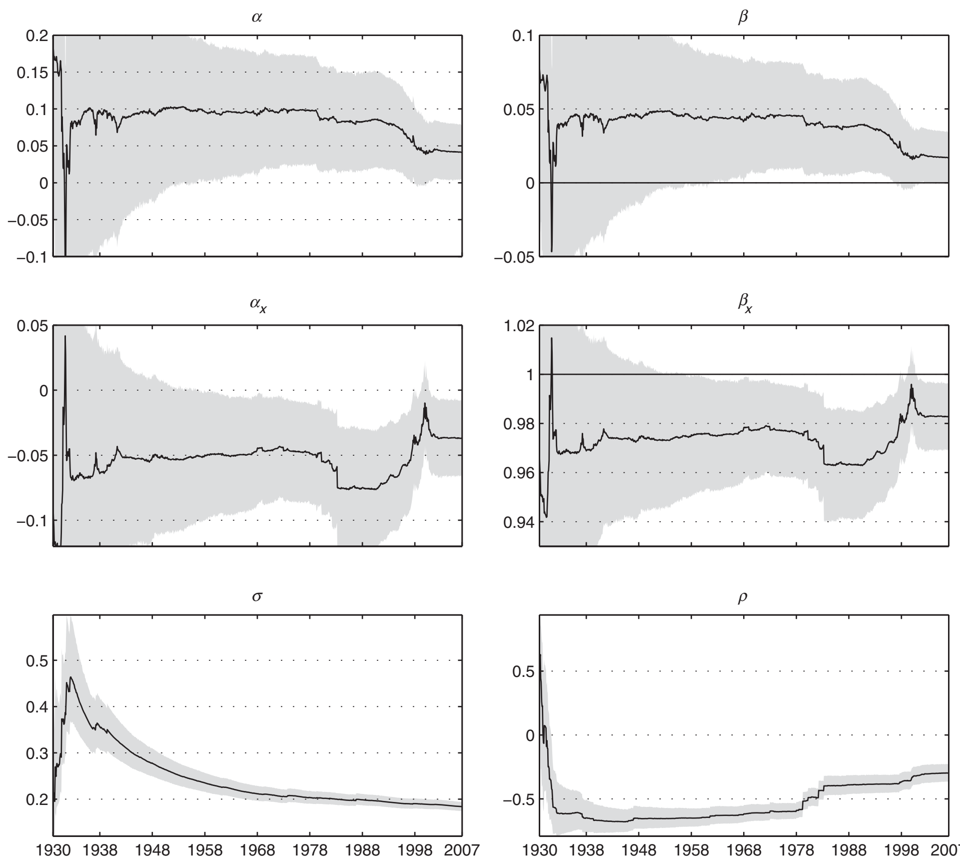

Figure 2. Sequential parameter estimates: CV model with net payout yield. This figure depicts sequential parameter estimates for the CV model

$$r_{t+1} = \alpha + \beta x_t + \sigma \varepsilon^r_{t+1}$$
$$x_{t+1} = \alpha_x + \beta_x x_t + \sigma_x \varepsilon^x_{t+1},$$

where r_{t+1} is the return on the market portfolio in excess of the risk-free rate from month t to month $t+1$. The predictor variable, x_t, is the net payout yield of Boudoukh et al. (2007). The shocks ε^r_{t+1} and ε^x_{t+1} are distributed standard normal with correlation coefficient ρ. For reference, the no-predictability line $\beta = 0$ is added to the plot for β, and the unit root line $\beta_x = 1$ is added to the plot for β_x. Each panel displays the posterior means and (1,99)% posterior probability intervals (the gray-shaded area) for each time period. Excess market return volatility, σ, is annualized.

the posterior mean (solid line) and a (1,99)% posterior probability interval at each point in time (the gray shaded area). The interval limits are not necessarily symmetric around the mean because the posteriors are exact finite-sample distributions.[13]

[13] Posterior probability intervals (also known as "credible intervals") represent the probability that a parameter falls within a given region of the parameter space, given the observed data. In

Three features in Figure 2 are worth noting. First, the speed of learning varies across parameters. Learning is far slower for expected return parameters, α and β, and parameters controlling the mean and speed of mean reversion of the dividend yield (α_x and β_x) than for the volatility and correlation parameters. Although standard asymptotics imply a common learning speed, there are differential learning speeds in finite samples. For the expected return parameters, there is still a significant amount of parameter uncertainty even after 30 or 40 years, highlighting the difficulty of learning expected return parameters due to the low signal-to-noise ratio and the persistence of the yield measure. The slow learning and substantial parameter uncertainty explains why estimation risk might be important for portfolio allocation.

Second, parameter estimates drift over time. This is especially true for the volatility parameters, whose drift occurs because the CV model has a constant volatility parameter, but is also true for the expected return parameters, as estimates of α and β decline slowly over the last 20 years of the sample. The estimates of β_x trend slightly upward, although the movement is not large.[14] This "drifting" of fixed parameter estimates is not necessarily surprising because the posterior distribution and posterior moments are martingales. Thus, the shocks to quantities such as $E(\alpha|y^t)$ are permanent and will not mean-revert.

Third, there is evidence of misspecification. For example, $E(\sigma|y^t)$ declines substantially over time, due to omitted stochastic volatility and the fact that the beginning of the sample has particularly high volatility. Since nearly all studies begin in 1927, discarding these data and starting postwar would create a serious sample selection bias. There are significant shifts in the mean parameters in the net payout yield equation, α_x and β_x, in the late 1970s and early 1980s. Interestingly, Boudoukh et al. (2007) formally test for a structural break and find no evidence, although we use monthly data, whereas they test using annual data. The source of the variation can be found in the time series of the regressors in Figure 1, where in the early 1980s the net payout variable has a series of high-frequency shocks. As discussed in the Internet Appendix, this is consistent with omitted stochastic volatility in the dividend yield process. The results from the other models are similar to those of the CV model, and are discussed in detail in the Internet Appendix.

One useful way to summarize the differences across models and regressors is to compare the predictability coefficients, that is, the β's in equation (2). Figure 3 shows that the estimated predictability coefficients differ across models for both data sets. The differences are quite large in the beginning of the sample, especially between the coefficients from constant models and those with stochastic volatility and time-varying regression coefficients. For the

Figure 2, the (1,99)% posterior probability interval represents the compact region of the parameter space for which there is a 1% probability that the parameter is higher than the region's upper bound and a 1% probability that it is lower than the lower bound. Posterior probability intervals should therefore not be interpreted in the same way as confidence intervals in classical statistics.

[14] One corroborating piece of evidence is in Brav et al. (2005), who present evidence that the speed of mean reversion for dividends has slowed in the last 50 years, making dividend yields more persistent.

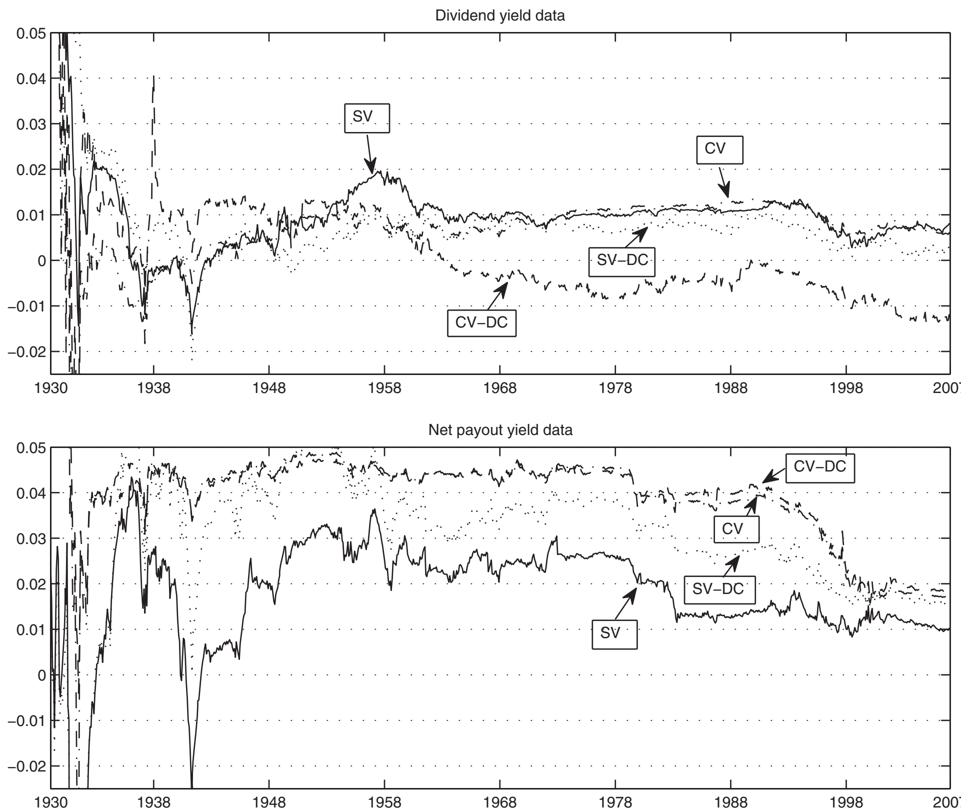

Figure 3. Predictability coefficient. This figure provides time series plots of the posterior mean of the predictability coefficient, β, across models and predictor variables. The top panel shows the coefficients of the four models using dividend yield as the predictor variable, and the bottom panel uses net payout yield as the predictor. CV and SV represent models with expected return predictability and constant volatility (CV) and stochastic volatility (SV), respectively. DC stands for drifting coefficients and represents models where the predictability coefficient is allowed to vary over time. For the DC models, we graph the loading on the predictor $\beta_0 + \beta_t$ from equation (6).

dividend yield data, the SV, SV-DC, and CV parameter estimates are quite similar in the latter part of the sample. For the net payout yield data, there are relatively large differences between the estimates over the entire sample period. The models with stochastic volatility have consistently lower coefficients than the models with constant volatility, with β coefficients almost half the size of the coefficients for the constant volatility models at certain points in the 1980s and 1990s. These differences are consistent with a time-varying signal-to-noise ratio. Overall, the models have varying degrees of statistical evidence in favor of return predictability. The Internet Appendix provides formal Bayesian hypothesis tests of predictability.

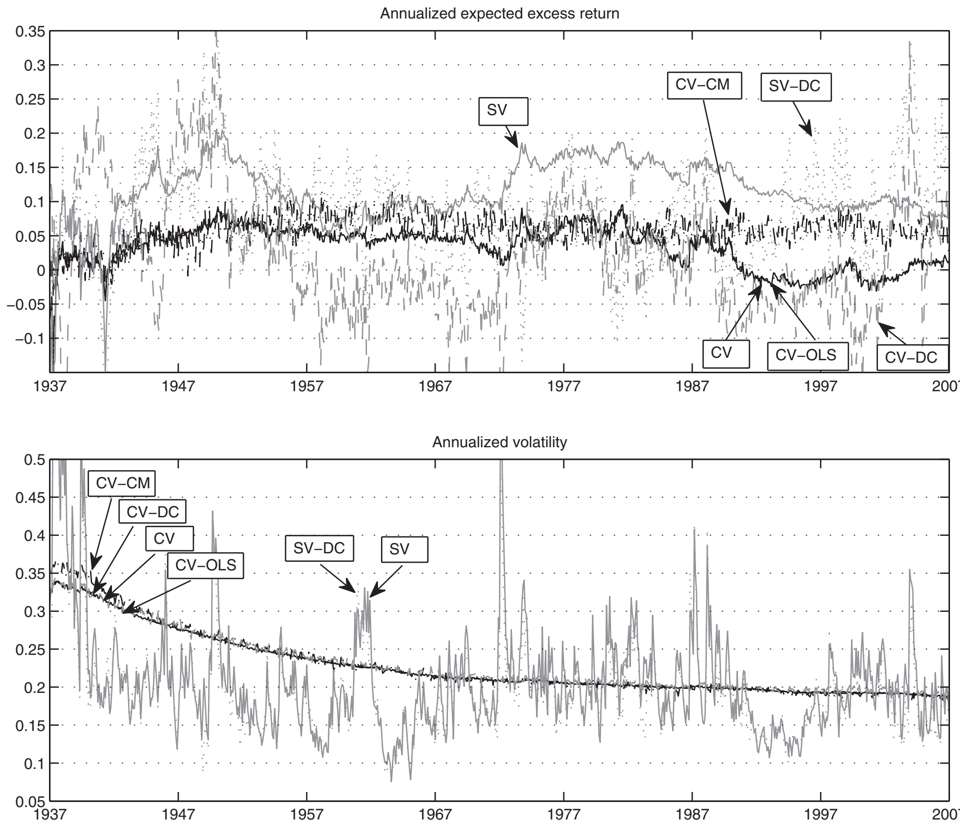

Figure 4. Predictive excess market returns. This figure provides time series plots of one-month-ahead predictive excess market returns across models, using dividend yield as the predictor variable. The top plot shows the annualized expected excess market return, and the bottom plot shows the annualized standard deviation.

A.2. Predictive Returns

Equation (13) provides the one-month-ahead predictive return distribution. To see the differences across models, Figure 4 plots predictive expected excess returns and volatilities for the dividend yield data set. A number of results are noteworthy. First, the models with expected return predictability have very different predictive expected returns due to the volatility specification. For example, the predictive expected returns in the SV model are higher than those in the CV model for nearly the entire sample, a result driven by the different signal-to-noise ratios in the models. Thus, the presence or absence of time-varying volatility impacts expected return estimates. Second, the drifting coefficient models generate very volatile expected return estimates, which is driven by the difficulty in estimating a time-varying regression coefficient with cash dividend yield, a weak signal. Finally, the predictive volatilities are dramatically different between the SV and CV specifications. While not surprising, this clearly shows the problems with CV specifications.

For each period, we also sequentially compute measures of fat tails, such as the predictive (conditional) kurtosis. The predictive distribution of the baseline CV model with constant volatility has an average (over the sample) excess kurtosis of 0.02, starting at about 0.15 and declining to less than 0.01 at the end of the sample. This slight excess kurtosis and its decline are due solely to parameter uncertainty, since there is no time-varying volatility in the CV model. Clearly, the models with constant volatility are incapable of generating any fat tails in the conditional distribution.

For the SV model, the average predictive excess kurtosis is 8.75, starting at approximately 15 in the beginning of the sample and declining to about 6 at the end of the sample. The initial higher kurtosis is due to the interaction between parameter uncertainty and stochastic volatility, as parameter uncertainty in the volatility equation fattens the tails of the volatility distribution, which in turn fattens the tails of predictive returns. This result is consistent with previous research showing that stochastic volatility models generate significant kurtosis at the monthly frequency (see Das and Sundaram (1999)). As mentioned earlier, the skewness of returns is modest and not statistically significant at monthly horizons. In the Internet Appendix, we provide additional results comparing the tail behavior of the models to those observed in the data. Overall, the stochastic volatility models are capable of generating more realistic tail behavior than the constant volatility models.

We also analyze the term structure of predictive volatilities. A provocative recent paper by Pástor and Stambaugh (2012) shows that predictive return volatility does not necessarily fall as the time horizon increases, in contrast to popular belief. Denoting $r_{t,t+k}$ as the return from time t to $t+k$, the authors find that $\text{var}(r_{t,t+k}|y^t)$ may increase as a function of k, due to parameter and state variable uncertainty. They document this feature in the context of a "predictive system" in which the relationship between the predictor variables and expected returns is imperfect but the conditional volatility of returns is constant.

We perform the same experiments as Pástor and Stambaugh (2012) for our model specifications, which are not formal imperfect predictive systems. The results, depicted in Figure 5, indicate that a number of our models generate increasing predictive volatility. Those with drifting coefficients are most similar to the results in Pástor and Stambaugh and have a striking increase in predictive volatility as the horizon increases. This is true of both CV and SV specifications with drifting coefficients. The SV model, when volatility is at its long-run mean, generates a slight upward slope in the predictive variance with parameter uncertainty, but a slight decrease conditional on fixed parameters. These results indicate that the increasing predictive volatility as a function of time horizon is a more general phenomenon, as it appears in models other than those considered in Pástor and Stambaugh (2012).

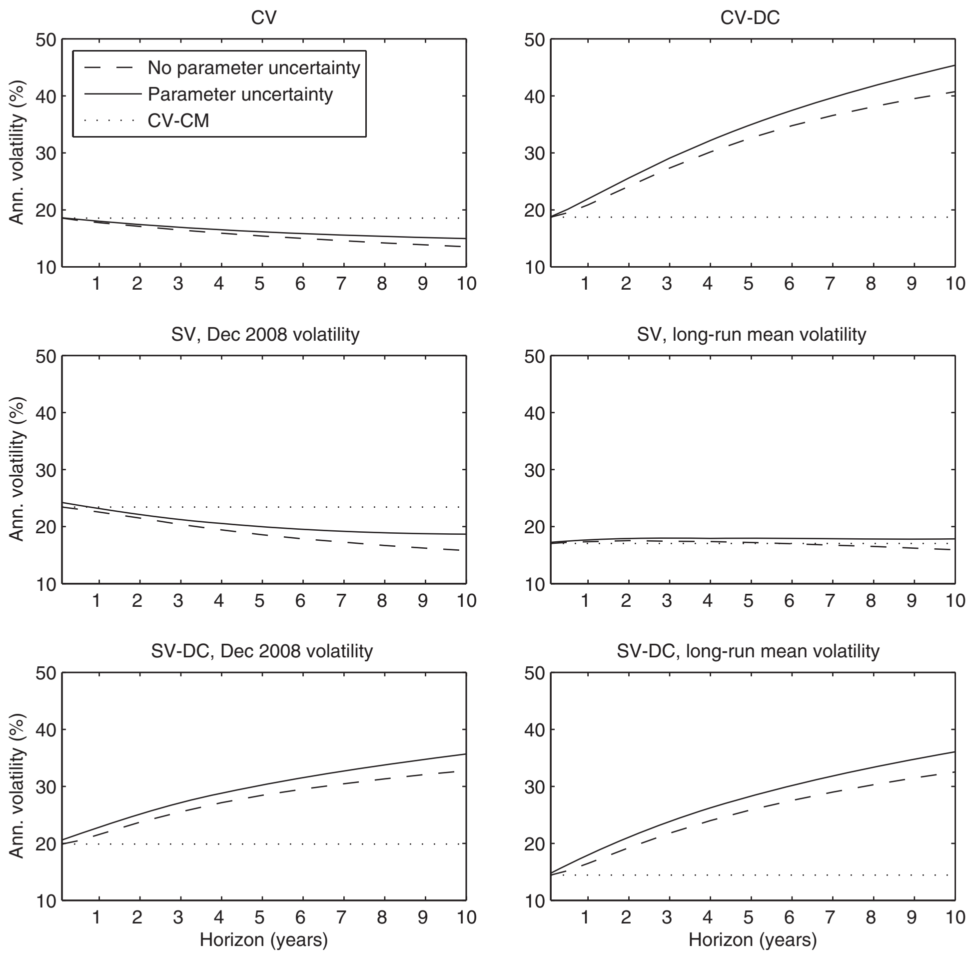

Figure 5. Predictive volatility. This graph shows the term structure of annualized predictive excess market return volatilities for various models, using the dividend yield as the predictor variable. Each panel represents a different model. The top-left plot shows the predictive volatility for the CV model, which has predictability in expected returns and constant variance. The striped line ignores parameter uncertainty, whereas the solid line includes the effect of parameter and state uncertainty. As a benchmark, the dotted horizontal line shows the volatility forecast from a CV-CM model (which has no predictability in either expected returns or variances) with the same current volatility estimate. The top-right plot shows the CV-DC model, with a drifting predictability coefficient and constant variance. The middle row of plots show the stochastic volatility (SV) model, which has both expected return and volatility predictability, at the realized volatility state at the last observation in our data set (December 2008, left plot), and at the average volatility level (right plot). Similarly, the bottom row shows the SV-DC model with stochastic volatility and a time-varying expected return predictability coefficient, both in December 2008 (left plot) and at the average level of volatility (right plot).

A.3. Model Comparison

As mentioned earlier, the particle filtering approach provides estimates of the cumulative marginal likelihood

$$p(y^t|\mathcal{M}_i) = \prod_{s=1}^{t-1} p(y_{s+1}|y^s, \mathcal{M}_i),$$

where

$$p(y_{t+1}|y^t, \mathcal{M}_i) = \int p\left(y_{t+1}|L_t, \theta, y^t, \mathcal{M}_j\right) p\left(L_t, \theta|y^t, \mathcal{M}_j\right) d\left(L_t, \theta\right). \qquad (15)$$

This accounts for all of the state and parameter uncertainty, and can be contrasted with standard maximum likelihood-based model comparison that conditions on parameter and state estimates. By integrating out the uncertainty in states and parameters, the Bayesian approach punishes needlessly complicated models and is often referred to as a "fully automatic Occam's razor" (Smith and Spiegelhalter (1980)).

Figure 6 reports posterior log-likelihoods, relative to the SV model, $\ln p(y^t|\mathcal{M}_i) - \ln p(y^t|\mathcal{M}_{SV})$, throughout the sample. This metric provides a relative comparison of how well an observation, y_t, conforms to its predictive distribution. Metrics that use the entire predictive distribution of returns are useful for data with time-varying volatility and nonnormalities. These metrics are fully out-of-sample, and the comparisons do not assume that one of the models is the "true" model, certainly a counterfactual assumption.

Three features stand out. First, both data sets eventually and strongly favor models with stochastic volatility. This is not a surprise, since time-varying volatility is such a strong feature of equity returns. Constant volatility models have two problems: they cannot capture the time variation in volatility and have very thin tails, as mentioned above. Second, different data sources favor different models. The cash dividend data suggest a SV model with constant dividend predictability, while net payout yield data favor a model with both stochastic volatility and drifting coefficients. This is intuitively sensible, since the better signal-to-noise ratio for expected returns when using net payout yield implies that it is possible to more accurately estimate the drifting coefficient and its underlying parameters. Below, we compare these results with the out-of-sample portfolio results, an alternative predictive measure of model performance.

Third, the shocks to the net payout yield in the early 1980s that are apparent in Figure 1 rapidly shift the relative likelihood toward models with stochastic volatility. These shocks happen around the enactment of SEC Rule 10b-18, which provided companies a safe harbor against share repurchase–related lawsuits, and induced more share repurchases going forward. Mechanically, the change in relative likelihoods occurs because these observations are ex ante unlikely under constant volatility, but far more likely under models with stochastic volatility.

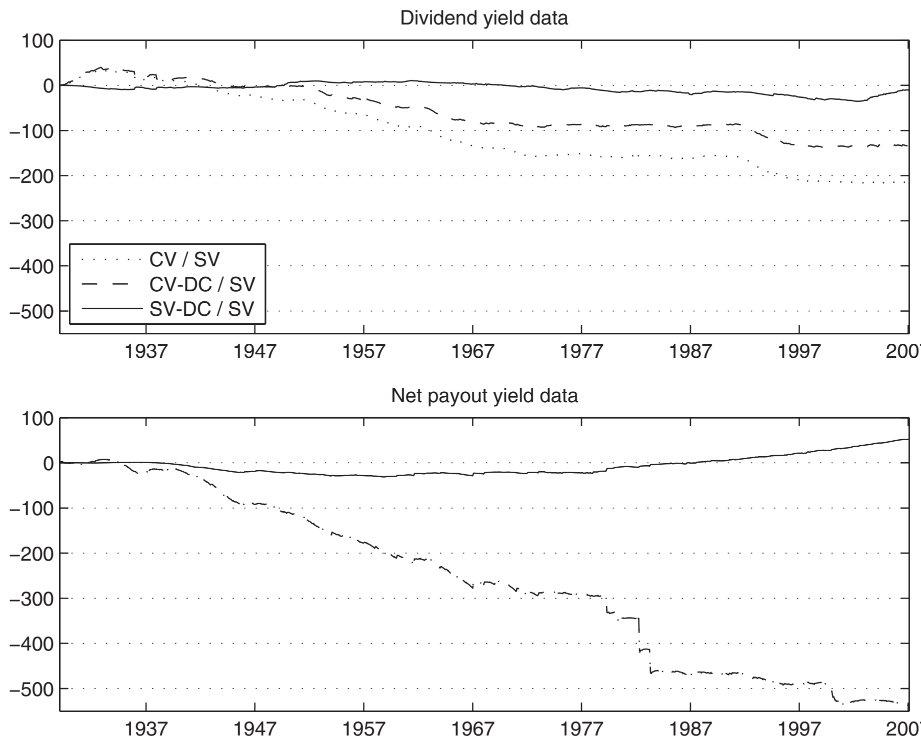

Figure 6. Relative log-likelihoods. This graph shows the time series of the log-likelihood of the CV, CV-DC, and SV-DC models, relative to the SV model. In each period, the likelihoods are calculated using all available data up to that period. The relative log-likelihoods for the dividend yield data are in the top panel, and for the net payout yield data in the bottom panel. CV and SV represent models with expected return predictability and constant volatility (CV) and stochastic volatility (SV), respectively. DC stands for drifting coefficients and represents models where the predictability coefficient is allowed to vary over time.

B. Portfolio Results

Tables I (cash dividends) and II (net payout yield) summarize the CEs and Sharpe ratios for the out-of-sample portfolio returns for each model, data set, risk aversion, and investment horizon. Table III reports the mean, standard deviation, skewness, and excess kurtosis statistic for out-of-sample returns for each time horizon and for $\gamma = 4$. We consider two null models to benchmark significance, a model with constant mean and variance and a model with constant mean and stochastic volatility.

B.1. Models Using Dividend Yields

Table I indicates that none of the constant volatility models provide statistically significant improvements in out-of-sample portfolio returns. Interestingly, among these poor performing models, the "best" model is actually

Table I

Portfolio Returns: Dividend Yield Data

This table shows out-of-sample portfolio returns using the dividend yield as the predictor for one-month, one-year, and two-year investment horizons. Investors have power utility with risk aversion parameter γ, and allocate their wealth between the market portfolio of stocks and a risk-free one-period bond. The certainty equivalent returns in Panel A represent the annualized risk-free return that gives the investor the same utility as the risky portfolio strategy. Panel B shows monthly Sharpe ratios. CM stands for a model with constant mean (i.e., no predictability), and CV and SV stand for constant and stochastic volatility, respectively. Hence, CV-CM represents a model with constant mean and constant volatility of returns. DC denotes drifting coefficients and represents models where the coefficient on net payout is allowed to vary over time. CV-OLS uses the OLS point estimates of equation (1), with data up to time t. CV-rolling OLS uses a 10-year rolling regression model to form portfolios. * and ** indicate that the result is significant at the 10% and 5% level, respectively, based on 500 simulated data sets with constant mean and volatility. † and †† indicate that the result is significant at the 10% and 5% level, respectively, based on 500 simulated data sets with constant mean and stochastic volatility.

	$\gamma = 4$			$\gamma = 6$		
	1m	1y	2y	1m	1y	2y
Panel A: Certainty Equivalent Returns (in % per annum)						
Constant volatility models						
CV-CM	4.77	5.05	5.05	4.38	4.61	4.61
CV-OLS	−8.60	1.14	−0.17	−25.22	1.70	−1.16
CV-rolling OLS	−19.68	−11.48	−12.59	−46.01	−24.71	−30.25
CV	−7.04	4.42	4.47	−2.47	4.09	4.03
CV-DC	−7.64	4.04	3.95	−3.29	3.72	3.51
Stochastic volatility models						
SV-CM	5.52	5.92	5.94	4.89	5.14	5.14
SV	6.43**,††	6.51**,††	6.53**,††	5.52**,††	5.64**,††	5.64**,††
SV-DC	6.56**,††	5.85	5.68	5.64**,††	5.17	5.02
Panel B: Sharpe Ratios (monthly)						
Constant volatility models						
CV-CM	0.089	0.099	0.099	0.088	0.101	0.101
CV-OLS	0.005	0.019	0.015	−0.001	0.016	0.006
CV-rolling OLS	0.102	0.108	0.110	0.100	0.108	0.109
CV	0.025	0.074	0.076	0.023	0.069	0.067
CV-DC	0.037	0.056	0.052	0.035	0.048	0.037
Stochastic volatility models						
SV-CM	0.132*,†	0.125	0.125	0.133*,†	0.124	0.124
SV	0.143**,††	0.144**,††	0.143**,††	0.143**,††	0.146**,††	0.145**,††
SV-DC	0.143**,††	0.135*	0.135*	0.144**,††	0.138*,†	0.136*

the CV-CM model, a model with constant mean and variance that accounts for parameter uncertainty. This model delivers CE yields of around 4.75% to 5% for $\gamma = 4$.[15] Models incorporating cash dividend yield as a regressor perform

[15] Note that the long-horizon CE yields in the CM models are different from those for the one-month horizon. The reason is that the monthly portfolio weights update each month whereas the long-horizon portfolio weights update annually.

Table II

Portfolio Returns: Net Payout Yield Data

This table shows out-of-sample portfolio returns using the net payout yield as the predictor for one-month, one-year, and two-year investment horizons. Investors have power utility with risk aversion parameter γ, and allocate their wealth between the market portfolio of stocks and a risk-free one-period bond. The certainty equivalent returns in Panel A represent the annualized risk-free return that gives the investor the same utility as the risky portfolio strategy. Panel B shows monthly Sharpe ratios. CM stands for a model with constant mean (i.e., no predictability), and CV and SV stand for constant and stochastic volatility, respectively. Hence, CV-CM represents a model with constant mean and constant volatility of returns. DC denotes drifting coefficients and represents models where the coefficient on net payout is allowed to vary over time. CV-OLS uses the OLS point estimates of equation (1), with data up to time t. CV-rolling OLS uses a 10-year rolling regression model to form portfolios. * and ** indicate that the result is significant at the 10% and 5% level, respectively, based on 500 simulated data sets with constant mean and volatility. † and †† indicate that the result is significant at the 10% and 5% level, respectively, based on 500 simulated data sets with constant mean and stochastic volatility.

	$\gamma = 4$			$\gamma = 6$		
	1m	1y	2y	1m	1y	2y
Panel A: Certainty Equivalent Returns (in % per annum)						
Constant volatility models						
CV-OLS	1.44	6.10	6.21	2.34	5.27	5.35
CV-rolling OLS	−14.85	−11.32	−11.23	−63.30	−61.95	−61.98
CV	3.03	6.07	6.19	3.29	5.25	5.33
CV-DC	0.77	6.19	6.27	2.00	5.33	5.40
Stochastic volatility models						
SV	6.85**,††	7.50**,††	7.67**,††	5.71**,††	6.29**,††	6.42**,††
SV-DC	6.23**,†	7.22**,††	7.36**,††	5.36**,†	6.08**,††	6.22**,††
Panel B: Sharpe Ratios (monthly)						
Constant volatility models						
CV-OLS	0.011	0.150**,††	0.151**,††	0.011	0.149**,††	0.151**,††
CV-rolling OLS	0.118	0.143**,††	0.146**,††	0.113	0.132*,†	0.136**,††
CV	0.045	0.155**,††	0.156**,††	0.045	0.154**,††	0.155**,††
CV-DC	0.044	0.156**,††	0.156**,††	0.042	0.155**,††	0.156**,††
Stochastic volatility models						
SV	0.155**,††	0.172**,††	0.172**††	0.154**,††	0.172**,††	0.171**,††
SV-DC	0.144**,††	0.166**,††	0.167**,††	0.145**,††	0.166**,††	0.166**,††

noticeably worse than the constant mean and volatility model. For example, at the one-month horizon, the CV model (constant volatility with cash dividend yield predictability) generates a CE yield of −7%. The two models ignoring parameter uncertainty, CV-OLS and CV-rolling OLS, perform particularly poorly, which quantifies the importance of parameter uncertainty, especially at short horizons. There is a modest improvement for some of the longer horizon cases, but they are not statistically significant.

These results are completely consistent with Goyal and Welch (2008). In fact, the results are even stronger as the CV model accounts for parameter

Table III
Portfolio Return Statistics

This table shows the first four moments (mean, standard deviation, skewness, and excess kurtosis) of annualized out-of-sample portfolio returns across one-month, one-year, and two-year investment horizons. Panel A reports the statistics for the portfolio where the investor uses the cash dividend yield as the predictor variable, and Panel B shows the statistics for the net payout yield as predictor. Investors have power utility with risk aversion parameter $\gamma = 4$, and allocate their wealth between the market portfolio of stocks and a risk-free one-period bond. The models are as described in Table I.

	One month				One year				Two year			
	Mean	St.dev.	Skew	Kurt	Mean	St.dev.	Skew	Kurt	Mean	St.dev.	Skew	Kurt
Panel A: Dividend Yield Data												
Constant volatility models												
CV-CM	0.059	0.070	−1.295	9.401	0.057	0.068	−0.714	4.452	0.057	0.068	−0.714	4.452
CV-OLS	0.036	0.139	−12.243	341.950	0.043	0.111	−5.680	74.737	0.043	0.115	−5.762	77.974
CV-rolling OLS	0.093	1.050	−4.518	47.891	0.145	0.349	−0.259	3.521	0.149	0.358	−0.308	3.409
CV	0.048	0.112	−7.848	193.460	0.051	0.068	−1.724	16.331	0.052	0.072	−1.778	17.452
CV-DC	0.047	0.123	−7.061	152.001	0.050	0.065	-1.750	11.763	0.048	0.064	−2.066	15.523
Stochastic volatility models												
SV-CM	0.104	0.154	−0.559	2.842	0.082	0.108	−0.488	2.562	0.082	0.108	−0.488	2.620
SV	0.095	0.119	0.123	4.349	0.075	0.079	0.107	3.276	0.077	0.085	0.134	3.588
SV-DC	0.099	0.138	−0.460	4.921	0.070	0.080	−0.678	4.333	0.069	0.077	−0.677	4.537
Panel B: Net Payout Yield Data												
Constant volatility models												
CV-OLS	0.040	0.084	−7.315	140.948	0.068	0.062	0.664	9.021	0.070	0.065	0.628	8.877
CV-rolling OLS	0.246	0.749	−1.637	19.032	0.194	0.316	−0.494	5.987	0.197	0.318	−0.490	5.809
CV	0.050	0.085	−4.219	67.174	0.067	0.057	0.733	8.623	0.069	0.060	0.676	8.369
CV-DC	0.052	0.103	−5.601	124.993	0.069	0.060	0.673	8.263	0.071	0.063	0.648	8.169
Stochastic volatility models												
SV	0.113	0.141	−0.777	8.935	0.092	0.093	0.101	3.589	0.097	0.101	0.120	3.452
SV-DC	0.108	0.143	−0.852	9.524	0.088	0.088	0.076	3.795	0.092	0.095	0.091	3.762

uncertainty, whereas Goyal and Welch do not. In every case, out-of-sample returns incorporating expected return predictability using the traditional cash dividend yield measure result in worse performance than a constant mean and volatility specification. Thus, even if you are Bayesian and account for parameter uncertainty, there is no statistical or economic evidence for out-of-sample gains for models with constant volatility.

The results in Panel A of Table III provide additional insights regarding the poor performance, showing that constant volatility models with cash dividend yields generate out-of-sample returns with extreme negative skewness and excess kurtosis. Intuitively, this occurs because the optimal portfolios in these models ignore fluctuating volatility (and therefore substantively fat-tailed return distributions), and thus the out-of-sample portfolio returns during high-volatility periods can be very large, leading to excess kurtosis. The CV-rolling OLS specification generates extremely high out-of-sample volatility, which generates the negative CE yields.

Turning to the stochastic volatility specifications, the SV-CM model incorporating time-varying volatility and a constant mean shows noticeably higher CEs and Sharpe ratios, but the increases are statistically insignificant in all cases except the short-horizon Sharpe ratios, which are significant at the 10% level. The effect of stochastic volatility is most clear from the portfolio return statistics in Table III. Compared to the CV-CM model, the SV-CM model has noticeable better skewness (-0.56 compared to -1.3) and lower excess kurtosis (2.84 compared to 9.40) for the one-month horizon. The longer horizon portfolio statistics also improve, though not by as much.

Portfolio returns in stochastic volatility models have lower kurtosis because persistent time-varying volatility tempers variation in portfolio volatility, as portfolio weights are lower in high-volatility periods and potentially increasing position in low-volatility environments. In addition, since large negative returns have historically occurred in periods of high volatility, the realized skewness of the portfolio returns is lower than the raw historical returns, since investors reduce their holdings in high-volatility periods. Thus, relative to the standard predictability model, stochastic volatility improves performance along all dimensions: higher Sharpe ratios, less negative skewness, and lower kurtosis.[16]

However, despite the improvements generated by the addition of time-varying volatility, the stochastic volatility portfolio returns are still not statistically significant relative to either of the null models. It is not easy to generate statistical significance against a benchmark of constant mean and variance in finite samples. This finding is important because it indicates that there is no statistical evidence for volatility timing, assuming a constant mean.

[16] Not surprisingly, as discussed in the Internet Appendix, the portfolio weights in the stochastic volatility models (for both dividend yield and net payout yield) are more negatively correlated with estimates of volatility than the constant volatility models. Since volatility is persistent, this partly explains a portion of the performance improvement in the stochastic volatility models. The stochastic volatility models also have higher average portfolio weights than the constant volatility models.

As mentioned earlier, there is no evidence in the literature for pure volatility timing over long time periods, although there is some evidence for a combination of volatility and correlation timing in the context of multiple asset portfolio problems (see Fleming, Kirby, and Ostdiek (2001, 2003)) subject to the caveats mentioned earlier.

Adding the dividend yield as a predictor, the full SV model (time-varying means and variances) does generate uniform statistically significant improvements, where significance is at the 5% level and holds against both null models. This is our first evidence that an ensemble of factors improves performance, and that there is an interaction between time-varying expected returns and time-varying volatility, as both features are needed to generate statistical significance. As discussed earlier, this is consistent with the importance of a time-varying signal-to-noise ratio for measuring time-varying expected returns.

Compared to the CV-CM model that also incorporates parameter uncertainty, the CEs in the SV model are more than 1.5% higher per year and the monthly Sharpe ratios increase from 0.089 to 0.143 for the $\gamma = 4$ case. Compared to the SV-CM model, predictable expected returns increase the CEs by about 1% per year and the Sharpe ratios increase from 0.132 to 0.143. The addition of expected return predictability using the dividend yield improves skewness and decreases volatility in a stochastic volatility model, with only a minor reduction in average return and a mild increase in kurtosis.

At all horizons, the returns generated by the SV model are always statistically significant against both null models. This uniform evidence is important to ensure the results are not specific to short-horizon problems. The portfolio returns generated by the full SV model generate a slight positive skew and only modest excess kurtosis. Estimation risk is also important in the SV model, as ignoring parameter and state uncertainty generally hurts out-of-sample performance.[17] Overall, we find strong evidence for statistically significant portfolio improvements at all horizons, using an ensemble of factors.

The models with drifting coefficients perform extremely poorly with constant volatility and better with stochastic volatility, but are only significant for short-horizon investors. This result should not be surprising, given the weak level of predictability and the additional parameters present in the DC models. As discussed earlier, with uncertain parameters, drifting coefficient models generate a strong increasing term structure of predictive volatility as in Pástor and Stambaugh (2012). We also note that both measures of model performance, the model probabilities and out-of-sample portfolios, identify the SV model as the best performing specification. Thus, the statistical and economic metrics coincide.

Some additional intuition for the relative performance can be seen in Figure 7, which shows the term structure of portfolio weights on a number

[17] The CE yields and Sharpe ratios for the SV model ignoring parameter and state uncertainty can decrease significantly. For example, for the case $\gamma = 4$, the one-month CE yields fall to 4.68% and 4.44% for the dividend yield and net payout measures, respectively, roughly 2% lower than the out-of-sample returns accounting for estimation risk (results not separately reported).

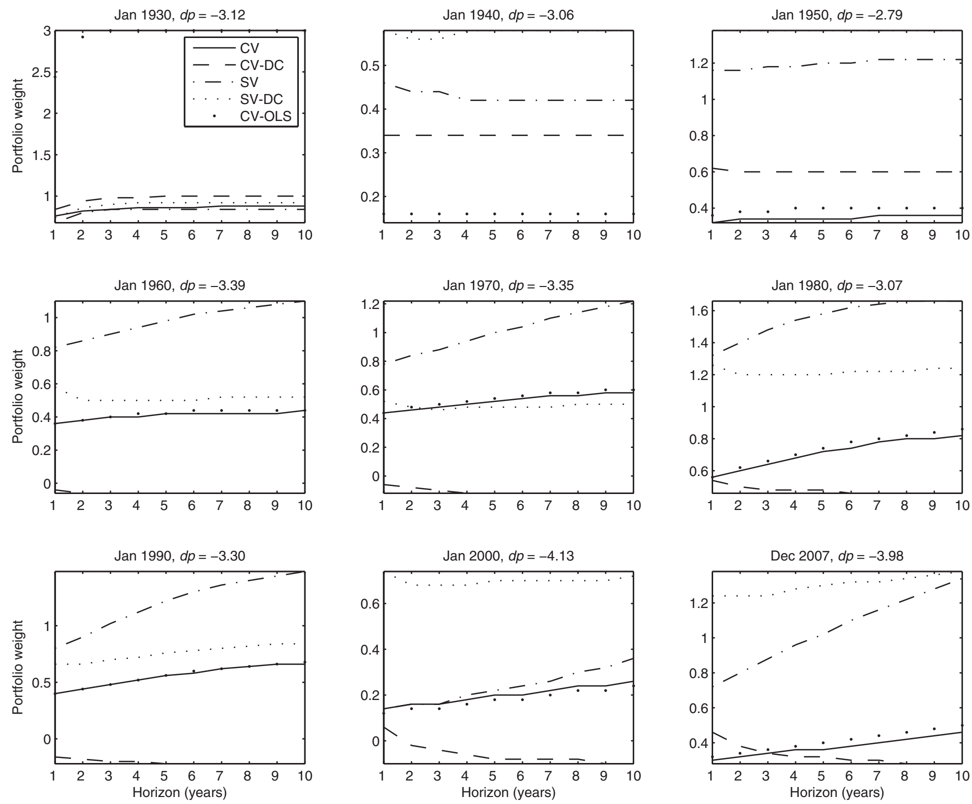

Figure 7. Optimal portfolio weights by investor horizon: Dividend yield data. This figure provides plots of optimal portfolio weights for an investor who allocates wealth between the market portfolio of stocks and a risk-free one-period bond, with an investment horizon spanning from 1 to 10 years. The plots show the optimal weights on the stock portfolio at the beginning of each decade in our sample period, as well as at the final data point in our sample (December 2008, bottom-right plot). The investor has power utility with risk aversion $\gamma = 4$, and rebalances annually while accounting for all parameter and state uncertainty. CV and SV represent models with expected return predictability and constant volatility (CV) and stochastic volatility (SV), respectively. DC stands for drifting coefficients and represents models where the predictability coefficient is allowed to vary over time. CV-OLS uses the OLS point estimates of equation (1), with data up to time t.

of different dates, for the cumulative OLS model (CV-OLS) and the Bayesian models, using the dividend yield as predictor. The different models generate very different long-horizon moments and return distributions, due to the time-varying state variables, estimation risk, and predictability. The differences arise because parameter uncertainty and mean-reversion (in expected returns and volatilities) impacts predictive moments differently as a function of investment horizon, as discussed above.

The key difference between the CV-OLS model and our CV-CM benchmark model is the effect of parameter uncertainty. At short horizons the CV-OLS and CV weights are quite similar, but at longer horizons the CV weight tends to be below the cumulative OLS weight, reflecting the fact that parameter uncertainty effectively increases the volatility of the predictive market returns (as seen in Figure 5). The difference tends to decrease as time progresses and more information arrives, reducing the importance of parameter uncertainty. We also see that the models with stochastic volatility tend to have higher portfolio weights. This happens because the initial high variance of the 1930s is only slowly unlearned by the constant variance models, resulting in lower portfolio weights for much of the sample period. This result underscores the importance of learning in a portfolio setting. The rolling regression portfolio weights are extremely variable and uninformative, with huge portfolio turnover, and for clarity we do not separately show them.[18]

B.2. Models Using Net Payout Yields

The portfolio results for the net payout predictor for each model are in Table II and Panel B of Table III. We find consistently statistically significant evidence for performance improvement for the SV specification using net payout yields, as the out-of-sample CE yields and Sharpe ratios are significant for every risk aversion and horizon combination. The statistical significance is relative to both null models, which indicates that it is the combination of time-varying volatility and expected return predictability that generates the significance.

In terms of economic significance, the CE yield for the short-horizon SV model portfolio is 6.85% (5.71%) for $\gamma = 4$ ($\gamma = 6$). This compares with a CE yield of 4.77% (4.38%) in the constant mean and variance case (from Table I). Thus, expected return and volatility timing increases CE returns by 1.5% to 2% per year, which, when compounded over a long sample, generates a dramatic and economically significant increase in realized utility. The monthly Sharpe ratio improves from 0.089 to roughly 0.155, an improvement of more than 70%.

The return statistics in Table II show strong improvement when compared to the CV specification at the one-month horizon, documenting the importance of time-varying volatility in controlling for tail behavior. Thus, there is strong statistical and economic evidence for the ability to time the investment set when using an ensemble of factors. Moreover, SV model performance using net payout yield is higher than that using dividend yield in every time horizon/risk aversion combination.

For the constant volatility models, none of the CE yields are statistically significant against either benchmark model, and the Sharpe ratio is insignificant at the monthly horizon for both risk aversions. For short-time horizons, the bottom panel in Table III shows that these models (excluding the CV-rolling OLS case) generate low average returns, extremely negative skewness, and very high kurtosis. The rolling regression case generates extremely high volatility. Thus, at the one-month horizon, none of the constant volatility models generate statistical significance.

[18] We report additional statistics regarding the portfolio weights in the Internet Appendix.

At longer horizons, many of the constant volatility specifications are statistically significant when performance is measured by the Sharpe ratio, but always insignificant when measured by the CE yield. This curious result can be reconciled by the skewness and kurtosis statistics, both of which are ignored when computing Sharpe ratios. For example, for the CV model, the skewness is −4.2 and kurtosis is 70. The CEs are insignificant for the constant volatility specifications because they take into account the higher moments due to the power utility specification. At longer horizons, the skewness improves for the constant volatility models, but the kurtosis is still generally greater than 10, which is penalized in the CE metric but not in the Sharpe ratio metric.

The drifting coefficients specification with time-varying volatility generates statistically significant gains in every risk aversion and investment horizon case. Even though the drifting coefficients model has three additional parameters, the increased signal-to-noise ratio of the net payout measure combined with stochastic volatility is sufficient to accurately estimate the drifting coefficient. Although significant across both metrics and for all time horizon/risk aversion cases, the SV model performs better in every case. Thus, there is a small, statistically insignificant loss from adding a drifting coefficients specification.

In terms of portfolio return statistics, Table III again documents the importance of stochastic volatility, as the kurtosis in the SV models is dramatically lower than in the constant volatility specifications. Assuming stochastic volatility and compared to the cash dividend yield case, net payout yields generate about 2% higher average returns and volatility per year, which translates into CE yields that are about 0.5% to 1% higher and Sharpe ratios that are approximately 0.1 to 0.3 higher. This improvement is due to the greater predictive ability of the net payout measure. Additional statistics on the portfolio weights are available in the Internet Appendix.

Overall, the results indicate that statistically and economically significant gains are generated for every time horizon and risk aversion case, provided that the investor (1) incorporates stochastic volatility, (2) incorporates expected return predictability using the net payout measure, and (3) accounts for parameter uncertainty when forming optimal portfolios. Together, this points toward the importance of an ensemble of factors needed to generate statistically significant out-of-sample portfolio improvements from predictability models. Expected return timing alone, even when accounting for parameter uncertainty, does not generate significant gains.

III. Conclusions

This paper studies the problem of an investor interested in forming optimal portfolios who learns about the investment set over time. To do so, we use various models that incorporate payout yield–based expected return predictors and stochastic volatility. The learning problem is Bayesian and is solved by using particle filters to sample from the posterior distribution of parameters, states, and models at each time period. After learning, our investor forms optimal portfolios by maximizing expected utility.

We reconcile seemingly contradictory evidence in the literature regarding the economic and statistical evidence for portfolio improvements generated by incorporating predictability. We find that an ensemble of factors capturing first-order important features of returns are needed to generate statistically significant portfolio improvements. In terms of models, it is important to incorporate stochastic volatility and time-varying expected returns, where the time variation in expected returns is captured by the net payout ratio from Boudoukh et al. (2007). It is also important to account for parameter uncertainty when forming optimal portfolios. We corroborate the findings in Goyal and Welch (2008) that simple predictability models with constant volatility do not lead to statistically significant out-of-sample portfolio gains, at least not for the set of predictors that we consider.

We also study the problem of sequential parameter inference and model monitoring, tracking relative model performance over time. We find strong time variation in the investor's beliefs about the parameters and model specifications. We find strong agreement between economic metrics of model performance (out-of-sample returns) and statistical metrics of model performance. We also connect our results and models to recent work by Pástor and Stambaugh (2009, 2012) on predictive systems and find that some of the specifications that we consider also have increasing volatility at longer horizons, even though the models are not necessarily predictive systems. These findings suggest that the Pástor and Stambaugh (2012) result may be far more general than their specific predictive systems model.

This analysis can be extended in a number of important directions. While we document statistically significant improvements in optimal portfolio performance, it should be possible to further improve performance by allowing for multiple predictor variables, developing more general model specifications, and incorporating economic restrictions as in Koijen and Van Binsbergen (2010). It would also be interesting to study optimal portfolios with alternative preferences that take into account a preference for early resolution of uncertainty, especially with model and parameter uncertainty.

Initial submission: June 2, 2010; Final version received: September 12, 2013
Editor: Campbell Harvey

REFERENCES

Amihud, Yakov, and Clifford M. Hurvich, 2004, Predictive regressions: A reduced-bias estimation method, *Journal of Financial and Quantitative Analysis* 39, 813–841.

Andersen, Torben G., Luca Benzoni, and Jesper Lund, 2002, An empirical investigation of continuous-time equity return models, *Journal of Finance* 57, 1239–1284.

Ang, Andrew, and Geert Bekaert, 2007, Stock return predictability: Is it there? *Review of Financial Studies* 20, 651–707.

Avramov, Doron, 2002, Stock return predictability and model uncertainty, *Journal of Financial Economics* 64, 423–458.

Bakshi, Gurdip, and Georgios Skoulakis, 2010, Do subjective expectations explain asset pricing puzzles? *Journal of Financial Economics* 98, 462–477.

Bandi, Federico M., Jeffrey R. Russell, and Yinghua Zhu, 2008, Using high-frequency data in dynamic portfolio choice, *Econometric Reviews* 27, 163–198.

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

Boudoukh, Jacob, Roni Michaely, Michael Roberts, and Matthew Richardson, 2007, On the importance of measuring payout yield: Implications for empirical asset pricing, *Journal of Finance* 62, 877–915.

Brandt, Michael W., Amit Goyal, Pedro Santa-Clara, and Jonathan R. Stroud, 2005, A simulation approach to dynamic portfolio choice with an application to learning about predictability, *Review of Financial Studies* 18, 831–873.

Brav, Alon, John R. Graham, Campbell R. Harvey, and Roni Michaely, 2005, Payout policy in the 21st century, *Journal of Financial Economics* 77, 483–527.

Brennan, Michael J., 1998, The role of learning in dynamic portfolio decisions, *European Finance Review* 1, 295–306.

Campbell, John Y., and Robert Shiller, 1988, Stock prices, earnings and expected dividends, *Journal of Finance* 43, 661–676.

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

Campbell, John Y., and Motohiro Yogo, 2006, Efficient tests of stock return predictability, *Journal of Financial Economics* 81, 27–60.

Carvalho, Carlos M., Michael S. Johannes, Hedibert F. Lopes, and Nicholas G. Polson, 2010, Particle learning and smoothing, *Statistical Science* 25, 88–106.

Carvalho, Carlos M., Michael S. Johannes, Hedibert F. Lopes, and Nicholas G. Polson, 2011, Particle learning for sequential Bayesian computation, *Bayesian Statistics* 9, 317–360.

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

Cremers, Martijn, 2002, Stock return predictability: A Bayesian model selection procedure, *Review of Financial Studies* 15, 1223–1249.

Dangl, Thomas, and Michael Halling, 2012, Predictive regressions with time-varying coefficients, *Journal of Financial Economics* 106, 157–181.

Das, Sanjiv R., and Rangarajan K. Sundaram, 1999, Of smiles and smirks: A term-structure perspective, *Journal of Financial and Quantitative Analysis* 34, 211–239.

Eraker, Bjorn, Michael S. Johannes, and Nicholas G. Polson, 2003, The impact of stochastic volatility and jumps in returns, *Journal of Finance* 58, 1269–1300.

Ferson, Wayne E., Sergei Sarkissian, and Timothy T. Simin, 2003, Spurious regressions in financial economics? *Journal of Finance* 53, 1393–1413.

Fleming, Jeff, Chris Kirby, and Barbara Ostdiek, 1998, Information and volatility linkages in the stock, bond, and money markets, *Journal of Financial Economics* 49, 111–137.

Fleming, Jeff, Chris Kirby, and Barbara Ostdiek, 2001, The economic value of volatility timing, *Journal of Finance* 56, 329–352.

Fleming, Jeff, Chris Kirby, and Barbara Ostdiek, 2003, The economic value of volatility timing using "realized" volatility, *Journal of Financial Economics* 67, 473–509.

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

Hansen, Lars, 2007, Beliefs, doubts and learning: Valuing macroeconomic risk, *American Economic Review* 97, 1–30.

Harvey, Campbell R., John C. Liechty, Merrill W. Liechty, and Peter Mueller, 2010, Portfolio selection with higher moments, *Quantitative Finance* 10, 469–485.

Harvey, Campbell R., and Akhtar Siddique, 2000, Conditional skewness in asset pricing tests, *Journal of Finance* 55, 1263–1295.

Henkel, Samuel J., J. Spencer Martin, and Federico Nardari, 2011, Time-varying short-horizon return predictability, *Journal of Financial Economics* 99, 560–580.

Hodrick, Robert J., 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *Review of Financial Studies* 5, 357–386.

Jacquier, Eric, Nicholas G. Polson, and Peter E. Rossi, 1994, Bayesian analysis of stochastic volatility models, *Journal of Business and Economic Statistics* 12, 371–389.

Jacquier, Eric, Nicholas G. Polson, and Peter E. Rossi, 2005, Bayesian analysis of correlated stochastic volatility models, *Journal of Econometrics* 122, 185–212.

Johannes, Michael S., and Nicholas G. Polson, 2006, Exact particle filtering and parameter learning, Working paper, University of Chicago.

Kandel, Shmuel, and Robert F. Stambaugh, 1996, On the predictability of stock returns: An asset allocation perspective, *Journal of Finance* 51, 385–424.

Kim, Sangjoon, Neil Shephard, and Siddhartha Chib, 1998, Stochastic volatility: Likelihood inference and comparison with ARCH models, *Review of Economic Studies* 65, 361–393.

Koijen, Ralph S.J., and Jules H. Van Binsbergen, 2010, Predictive regressions: A present-value approach, *Journal of Finance* 65, 1439–1471.

Lettau, Martin, and Sydney Ludvigson, 2001, Consumption, aggregate wealth, and expected stock returns, *Journal of Finance* 56, 815–850.

Lettau, Martin, and Stijn van Nieuwerburgh, 2008, Reconciling the return predictability evidence, *Review of Financial Studies* 21, 1607–1652.

Lewellen, Jonathan, 2004, Predicting returns with financial ratios, *Journal of Financial Economics* 74, 209–235.

Menzly, Lior, Tano Santos, and Pietro Veronesi, 2004, Understanding predictability, *Journal of Political Economy* 112, 1–47.

Pástor, Ľuboš, and Robert F. Stambaugh, 2009, Predictive systems: Living with imperfect predictors, *Journal of Finance* 64, 1583–1628.

Pástor, Ľuboš, and Robert F. Stambaugh, 2012, Are stocks really less volatile in the long run? *Journal of Finance* 67, 431–478.

Paye, Bradley S., and Allan Timmermann, 2006, Instability of return prediction models, *Journal of Empirical Finance* 13, 274–315.

Rosenberg, Barr, 1972, The behavior of random variables with nonstationary variance and the distribution of security prices, Working paper, University of California, Berkeley.

Santos, Tano, and Pietro Veronesi, 2006, Labor income and predictable stock returns, *Review of Financial Studies* 19, 1–44.

Shanken, Jay A., and Ane Tamayo, 2012, Dividend yield, risk and mispricing: A Bayesian analysis, *Journal of Financial Economics* 105, 131–152.

Shiller, Robert J., 1981, Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review* 71, 421–436.

Smith, Adrian F.M., and David J. Spiegelhalter, 1980, Bayes factors and choice criteria for linear models, *Journal of the Royal Statistical Society Series B* 42, 213–220.

Stambaugh, Robert F., 1986, Bias in regressions with lagged stochastic regressors, CRSP working paper 156, University of Chicago.

Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.

Torous, Walter, Rossen Valkanov, and Shu Yan, 2004, On predicting stock returns with nearly integrated explanatory variables, *Journal of Business* 77, 937–966.

Wachter, Jessica A., and Missaka Warusawitharana, 2009, Predictable returns and asset allocation: Should a skeptical investor time the market? *Journal of Econometrics* 148, 162–178.

Wachter, Jessica A., and Missaka Warusawitharana, 2012, What is the chance that the equity premium varies over time? Evidence from predictive regressions, NBER working paper 17334, University of Pennsylvania Wharton School of Business.

Xia, Yihong, 2001, Learning about predictability: The effect of parameter learning on dynamic asset allocation, *Journal of Finance* 56, 205–246.

Yan, Jun, 2005, Asymmetry, fat-tail, and autoregressive conditional density in financial return data with systems of frequency curves, Working paper, University of Iowa.