OXFORD

Option-Implied Equity Premium Predictions via Entropic Tilting

Konstantinos Metaxoglou[1], Davide Pettenuzzo[2] and Aaron Smith[3]

[1]Carleton University, [2]Brandeis University and [3]University of California, Davis

Address correspondence to D. Pettenuzzo Department of Economics, Brandeis University 415 South Street, MS 021 Waltham, MA 02453-2728, USA, or e-mail: dpettenu@brandeis.edu.

Abstract

We propose a new method to improve density forecasts of the equity premium using information from options markets. We obtain predictive densities from stochastic volatility (SV) and GARCH models, which we then tilt using the second moment of the risk-neutral distribution implied by options prices while imposing a non-negativity constraint on the equity premium. By combining the backward-looking information contained in the GARCH and SV models with the forward-looking information from options prices, our procedure improves the performance of predictive densities. Using density forecasts of the U.S. equity premium from January 1990 to December 2014, we find that tilting leads to more accurate predictions using statistical and economic criteria.

Key words: density forecasts, entropic tilting, equity premium, options, variance risk premium

JEL classification: C11, C22, G11, G12

Empirical asset pricing usually employs forecasting models that are *backward looking*—they use past observations on a set of variables to project future asset returns. The set of variables is often motivated by economic theory—for example, macroeconomic and financial variables, such as the dividend yield or the term spread. On the other hand, derivative prices convey information about the conditional density of future outcomes and hence are inherently *forward looking*. They contain information about market expectations and thus, should be useful for improving return forecasts.

In this paper, we provide a simple procedure to blend backward- and forward-looking information to refine the predictive densities of the equity premium obtained from a baseline econometric model. Our approach entails taking a given predictive density for excess returns and tilting it using moments implied by options prices. Specifically,

we proceed by extracting the variance of the risk-neutral distribution of returns from options prices, and subtracting from it a regression-based estimate of the variance risk premium to obtain a forward-looking variance estimate. In the spirit of Robertson, Tallman, and Whiteman (2005), we then rely on entropic tilting to twist the original predictive distribution using this forward-looking variance, while at the same time imposing a non-negativity constraint on the first moment. The latter constraint has been shown to substantially improve the out-of-sample (OOS) predictability of excess returns; see Campbell and Thompson (2008) and, more recently, Pettenuzzo, Timmermann, and Valkanov (2014).

Our procedure is simple and has a low computational cost; using a few lines of code, we modify the original predictive density such that its moments conform with the additional restrictions we wish to impose. We apply our method to S&P500 returns using either a stochastic volatility (SV) or a GARCH model to form the baseline predictive density. In both cases, we find that tilting the baseline density using our procedure significantly improves the OOS predictability of stock returns in terms of statistical and economic measures of forecasting accuracy.

Our paper contributes to a rapidly growing literature that looks at the role of option-implied information in improving forecasts. In particular, several papers show that option-implied volatility can predict future realized volatility as well as the equity premium; see, for example, Szakmary et al. (2003) and Bollerslev, Tauchen, and Zhou (2009). We make two contributions to this literature. First, we provide a highly flexible non-parametric method for incorporating option-implied moments into baseline forecasts. Second, we work with density forecasts, whereas the bulk of the existing literature incorporates option-implied moments among the predictors in a point forecasting regression; see Altigan, Bali, and Demitras (2015) for a recent example. Finally, it is worth noting that our method can be easily extended to higher moments, such as skewness and kurtosis, which have received increased attention recently in empirical asset pricing (Young Chang, Christoffersen, and Jacobs, 2013).

The remainder of the paper is organized as follows. In Section 1, we describe the entropic tilting procedure, along with our approach to constructing the model-based predictive densities for the equity premium. Our approach for removing the variance risk premium from the variance of the risk-neutral distribution implied by options prices follows. Section 2 presents our main results and Section 3 focuses on OOS statistical and economic performance. We present some robustness checks in Section 4. Finally, Section 5 provides some concluding remarks.

1 Entropic Tilting for Equity Premium Forecasting

Entropic tilting is a highly flexible non-parametric method to change the shape of a distribution to incorporate additional information about a random variable of interest. Such additional information may come in the form of moments and this is the approach we take here. In what follows, we start from the predictive density of the excess returns on the S&P500 implied by either a SV or a GARCH model. We then use entropic tilting to alter this baseline distribution to incorporate moment restrictions derived from options prices and economic theory.

We begin by first outlining the general entropic tilting method and our approach to incorporating the moment-based information from the options markets into a baseline predictive density. Next, we describe the econometric model we use to produce the baseline density forecasts. We conclude this section by describing our approach to removing the variance risk premium from the risk-neutral variance we derive from options prices.

1.1 General Method

Let $p(r_{t+1}|\mathcal{D}^t)$ denote the baseline predictive density for the equity premium r_{t+1} with $\mathcal{D}^t$ being the information set available at time t, and $t = 1, ..., T - 1$. The econometrician is assumed to have additional information about a function $g(r_{t+1})$, which was not used to generate the baseline predictive density. This additional information takes the form of moments of $g(r_{t+1})$ such that

$$E[g(r_{t+1})|\mathcal{D}^t] = \overline{g}_t. \tag{1}$$

For example, $g(r_{t+1})$ may represent quantities such as the mean, $g(r_{t+1}) = r_{t+1}$, the variance, $g(r_{t+1}) = (r_{t+1} - E[r_{t+1}|\mathcal{D}^t])^2$, or higher moments of the predictive distribution; see Robertson, Tallman, and Whiteman (2005) for a very informative exposition. The information could be in the form of moment restrictions implied by economic theory, such as Euler conditions in Giacomini and Ragusa (2014), or could be coming from survey forecasts and model-based nowcasts as in Altavilla, Giacomini, and Constantini (2014) and Krüger, Clark, and Ravazzolo (2015).

Generally, the expected value of $g(r_{t+1})$ under the baseline distribution will not equal $\overline{g}_t$

$$\int g(r_{t+1})p(r_{t+1}|\mathcal{D}^t)\mathrm{d}r_{t+1} \neq \overline{g}_t. \tag{2}$$

Thus, by transforming $p(r_{t+1}|\mathcal{D}^t)$ so that Equation (1) holds, we refine the baseline predictive density. To implement the method, consider N random draws from the baseline predictive distribution $p(r_{t+1}|\mathcal{D}^t)$. We denote these draws with $\{r_{t+1}^i\}_{i=1}^N$, where each draw is associated with a weight $\pi_i = 1/N$. We construct a new set of weights $\{\pi_{it}^*\}_{i=1}^N$ that represent a new predictive density that is as close as possible to the baseline and also satisfies the moment restriction implied by Equation (1). Following a standard approach in the literature, we use the empirical Kullback–Leibler Information Criterion (KLIC) to measure the distance between the baseline and the new predictive density[1]

$$\mathrm{KLIC}(\pi_t^*; \pi) = \sum_{i=1}^N \pi_{it}^* \ln\left(\frac{\pi_{it}^*}{\pi_i}\right). \tag{3}$$

The objective is to find new weights that minimize Equation (3) subject to the constraints

$$\pi_{it}^* \geq 0, \quad \sum_{i=1}^N \pi_{it}^* = 1, \quad \sum_{i=1}^N \pi_{it}^* g(r_{t+1}^i) = \overline{g}_t, \tag{4}$$

---

[1] Other measures of divergence are also available. As Giacomini and Ragusa (2014) note, the KLIC provides a convenient analytical expression for the tilted weights and, unlike other measures of distance, it has a direct counterpart in the logarithmic scoring rule, which is a common and well-studied measure for evaluating density forecasts (Amisano and Giacomini, 2007).

where the last constraint may be viewed as the Monte-Carlo approximation to the moment restriction in Equation (1) using the language in Cogley, Morozov, and Sargent (2005).[2] The implied first-order conditions are given by

$$1 + \ln\left(\frac{\pi_i}{\pi_{it}^*}\right) - \mu_t - \gamma_t' g(r_{t+1}^i) = 0, \quad i = 1, \dots, N \tag{5}$$

with μ_t and γ_t being the Lagrange multipliers associated with the adding-up and moment constraints. The new weights are then given by

$$\pi_{it}^* = \frac{\pi_i \, \exp\left(\gamma_t^{*'} g(r_{t+1}^i)\right)}{\sum_{i=1}^N \pi_i \, \exp\left(\gamma_t^{*'} g(r_{t+1}^i)\right)}. \tag{6}$$

As a result, the baseline weights are *tilted* in an exponential fashion via Equation (6) to generate the new weights. The tilting parameter γ_t^* can be found by solving the minimization problem

$$\gamma_t^* = \arg \min_{\gamma_t} \sum_{i=1}^N \exp\left(\gamma_t'\left[g(r_{t+1}^i) - \overline{g}_t\right]\right). \tag{7}$$

In our case, we use the variance of the risk-neutral distribution for the equity premium, as implied by the options markets, to distort the baseline predictive distribution $p(r_{t+1}|\mathcal{D}^t)$ so that its dispersion, as captured by $\mathrm{Var}(r_{t+1}|\mathcal{D}^t)$, resembles that of the option-implied risk-neutral distribution. It is the forward-looking aspect of the options markets that serves as the source of new information and is also the novelty in our approach. In addition, we follow the recent literature on stock return predictability (e.g., Campbell and Thompson, 2008; Pettenuzzo, Timmermann, and Valkanov, 2014) and further impose a non-negativity constraint on the first moment of the tilted predictive density. In the spirit of Robertson et al., we incorporate restrictions, which could be built directly into the forecasting model, in a manner that is less demanding computationally.[3]

1.2 Baseline Predictive Densities

There is ample evidence pointing to time variation in both the conditional mean and volatility of the return distribution; see, for example, Rapach and Zhou (2013) and Andersen et al. (2006). In this section, we discuss two approaches for obtaining baseline predictive densities for returns. The first is Bayesian and is based on a SV model estimated using a Gibbs sampler. The second is frequentist and is based on maximum-likelihood estimation (MLE) of a GARCH(1,1) model. Although Bayesian estimation of GARCH models is possible (e.g., Bauwens and Lubrano, 1998; Chan and Grant, 2016), we want to show that our method can be easily implemented with a standard estimation approach and proceed with MLE.

---

2 See Robertson et al. and the references therein. The 2012 *Econometric Reviews* Special Issue on Entropy and the 2002 *Journal of Econometrics* Issue on Information and Entropy Econometrics offer more detailed treatments on entropy and the use of alternative divergence measures.

3 Our approach to handle the inequality constraint on the expected return regressions follows very closely the implementation of Campbell and Thompson (2008). That is, we use the entropic tilting procedure to shift the posterior mean of the predictive densities toward zero only when the predictive densities are centered around a negative value, without changing their first moment—therefore, only altering their second moment—in all other instances.

1.2.1 Stochastic volatility

We rely on the following model with time-varying first and second moments to produce the baseline predictive density $p(r_{t+1}|\mathcal{D}^t)$ of the monthly equity premium

$$r_{\tau+1} = \mu + \beta' x_\tau + \exp\left(h_{\tau+1}\right) u_{\tau+1}, \quad \tau = 1, ..., t-1, \tag{8}$$

where $h_{\tau+1}$ represents the log-volatility at time $\tau + 1$, x_τ denotes a (vector of) lagged predictor(s), and $u_{\tau+1} \sim \mathcal{N}(0, 1)$. We further assume that the log-volatility $h_{\tau+1}$ follows an autoregressive process and depends on lagged intra-month information in the form of realized volatility

$$h_{\tau+1} = \lambda_0 + \lambda_1 h_\tau + \lambda_2 \mathrm{RV}_\tau + \xi_{\tau+1}, \quad \xi_{\tau+1} \sim \mathcal{N}\left(0, \sigma_\xi^2\right), \tag{9}$$

where RV_τ denotes the realized volatility at time τ, computed by summing the squared daily returns within month τ, and $|\lambda_1| < 1$.[4] Note also that u_τ and ξ_s are mutually independent for all τ and s.

We estimate the parameters in Equation (8) using Bayesian methods. Following standard practice in the Bayesian literature (Koop, 2003), the priors for μ and β in Equation (8) are assumed to be normal

$$\begin{bmatrix} \mu \\ \beta \end{bmatrix} \sim N(\underline{b}, \underline{V}). \tag{10}$$

For the hyperparameters $\underline{b}$ and $\underline{V}$, we set aside an initial training sample of t_0 observations to calibrate them (e.g., Primiceri, 2005; Clark, 2011) and proceed as follows:

$$\underline{b} = \begin{bmatrix} \bar{r}_{t_0} \\ 0 \end{bmatrix}, \quad \underline{V} = \underline{\psi}^2 \left[s_{r,t_0}^2 \left(\sum_{\tau=1}^{t_0-1} x_\tau x_\tau' \right)^{-1} \right], \tag{11}$$

where

$$\bar{r}_{t_0} = \frac{1}{t_0 - 1} \sum_{\tau=1}^{t_0-1} r_{\tau+1}, \quad s_{r,t_0}^2 = \frac{1}{t_0 - 2} \sum_{\tau=1}^{t_0-1} \left(r_{\tau+1} - \bar{r}_{t_0}\right)^2.$$

Our choice of $\underline{b}$ in Equation (11) reflects the prior belief that the best predictor of stock returns is the average of past returns. Therefore, we center the prior intercept on the historical average (HA) of the excess returns while we set the prior mean on the slope coefficient(s) to zero. Furthermore, the scalar $\underline{\psi}$ in Equation (11) controls the tightness of the prior ($\underline{\psi} \to \infty$ corresponds to a diffuse prior on μ and β). We specify rather uninformative priors and set $\underline{\psi} = 1.0e^6$.

We also require priors on the sequence of volatilities, $h^t = \{h_1, ..., h_t\}$, and the SV parameters λ_0, λ_1, λ_2, and σ_ξ^2. Decomposing the joint probability of these parameters and using Equation (9), we have

$$p\left(h^t, \lambda_0, \lambda_1, \lambda_2, \sigma_\xi^{-2}\right) = p\left(h^t|\lambda_0, \lambda_1, \lambda_2, \sigma_\xi^{-2}\right) p(\lambda_0, \lambda_1, \lambda_2) p\left(\sigma_\xi^{-2}\right)$$

$$= \left[\prod_{\tau=1}^{t-1} p\left(h_{\tau+1}|\lambda_0, \lambda_1, \lambda_2, h_\tau, \sigma_\xi^{-2}\right) p(h_1) \right] p(\lambda_0, \lambda_1, \lambda_2) p\left(\sigma_\xi^{-2}\right), \tag{12}$$

---

4 We restrict $|\lambda_1| < 1$ to ensure that volatility is stationary and mean-reverting around RV_τ.

where

$$h_{\tau+1}|\lambda_0, \lambda_1, \lambda_2, h_\tau, \sigma_\xi^{-2} \sim \mathcal{N}\left(\lambda_0 + \lambda_1 h_\tau + \lambda_2 \mathrm{RV}_\tau, \sigma_\xi^2\right), \qquad \tau = 1, ..., t-1. \tag{13}$$

To complete the prior elicitation for $p\left(h^t, \lambda_0, \lambda_1, \lambda_2, \sigma_\xi^{-2}\right)$, we choose priors for $\lambda_0, \lambda_1, \lambda_2$, the initial log volatility h_1, and σ_ξ^{-2}, from the normal-gamma family

$$h_1 \sim \mathcal{N}\left(\ln\left(s_{r,t_0}\right), \underline{k}_h\right), \tag{14}$$

$$\begin{bmatrix} \lambda_0 \\ \lambda_1 \\ \lambda_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \underline{m}_{\lambda_0} \\ \underline{m}_{\lambda_1} \\ \underline{m}_{\lambda_2} \end{bmatrix}, \begin{bmatrix} \underline{V}_{\lambda_0} & 0 & 0 \\ 0 & \underline{V}_{\lambda_1} & 0 \\ 0 & 0 & \underline{V}_{\lambda_2} \end{bmatrix} \right), \quad \lambda_1 \in (-1, 1), \tag{15}$$

and

$$\sigma_\xi^{-2} \sim \mathcal{G}\left(1/\underline{k}_\xi, \underline{v}_\xi(t_0 - 1)\right). \tag{16}$$

We set $\underline{k}_\xi = 0.5$, $\underline{v}_\xi = 10$, and $\underline{k}_h = 10$. These choices restrict changes to the log-volatility to be roughly equal to 0.7, on average, and place a relatively diffuse prior on the initial log-volatility state.

Following Clark and Ravazzolo (2015), the hyperparameters are as follows: $\underline{m}_{\lambda_0} = \underline{m}_{\lambda_3} = 0$, $\underline{m}_{\lambda_1} = 0.9$, $\underline{V}_{\lambda_0} = \underline{V}_{\lambda_3} = 0.25$, and $\underline{V}_{\lambda_0} = 1.0\mathrm{e}^{-4}$. This corresponds to setting the prior means and standard deviations for the intercept and RV coefficient to 0 and 0.5, respectively. As for the AR(1) coefficient, these choices imply a prior mean of 0.9 with a standard deviation of 0.01. Overall, these are informative priors that match the persistent dynamics in the log volatility process.

We estimate the model in Equations (8) and (9) using a Gibbs sampler that lets us compute posterior draws for μ, β, h^t, σ_ξ^{-2}, λ_0, λ_1, and λ_2. These draws are used to compute density forecasts for r_{t+1}

$$\begin{aligned} p(r_{t+1}|\mathcal{D}^t) = \int & p(r_{t+1}|h_{t+1}, \Theta, h^t, \mathcal{D}^t) \\ & \times p(h_{t+1}|\Theta, h^t, \mathcal{D}^t) p(\Theta, h^t|\mathcal{D}^t) \mathrm{d}\Theta \mathrm{d}h^{t+1}. \end{aligned} \tag{17}$$

where $\Theta = \left(\mu, \beta, \sigma_\xi^{-2}, \lambda_0, \lambda_1, \lambda_2\right)$ contains the time-invariant parameters.[5] Section A of the Online Appendix contains details on the Gibbs sampler and the computation of the integral in Equation (17).

1.2.2 GARCH
Our second approach to generate baseline density forecasts is based on a GARCH(1,1) model. Although there is a plethora of GARCH-type models with well-documented properties, a GARCH(1,1) model with only three parameters in the conditional volatility equation

---

5　Throughout the paper, we run the Gibbs samplers for a total of 25,000 iterations, after a first set of 2500 draws is discarded to allow the samplers to achieve convergence. We further thin the MCMC chains by keeping one out of every five draws. This yields a total of 5000 retained draws for each model and time period within the forecast evaluation window.

is very often adequate to obtain a reasonably good fit for financial time series (Zivot, 2009). Hence, we proceed with the following setup

$$r_{\tau+1} = \mu + \beta' x_\tau + h_{\tau+1}^{1/2} u_{\tau+1}, \quad \tau = 1, ..., t-1, \tag{18}$$

$$h_{\tau+1} = \lambda_0 + \lambda_1 h_\tau + \lambda_2 h_\tau u_\tau^2, \quad u_\tau \sim \mathcal{N}(0, 1). \tag{19}$$

We then use the following two-step approach to generate the baseline density forecasts. In the first step, we obtain recursive maximum-likelihood estimates of the parameters in the mean and volatility equations, $\hat{\theta}_{G\tau} = \left(\hat{\mu}, \hat{\beta}' \hat{\lambda}_0, \hat{\lambda}_1, \hat{\lambda}_2 \right)'$, and the associated variance–covariance matrix, $\hat{V}_{G\tau}$, using an expanding-window approach. In the second step, we generate a large number of one-period ahead return predictions by plugging draws from $\mathcal{N}\left(\hat{\theta}_{G\tau}, \hat{V}_{G\tau} \right)$ into Equations (18) and (19) to obtain the appropriate time-varying mean and volatility of the return distribution to generate the GARCH analog of $p(r_{t+1}|\mathcal{D}^t)$.

1.3 Removing the Variance Risk Premium

We capitalize on the literature that has demonstrated the predictive power of implied volatility for future realized volatility; see Jorion (1995) and, more recently, Szakmary et al. (2003), among others. The basic argument is that implied volatility—inferred from options data as in our case—can be perceived as the market's expectation of future volatility and, hence, it is a market-based volatility forecast (Poon and Granger, 2003). The feature of the implied volatility that is particularly appealing for a forecasting exercise like the one undertaken here is that it is inherently forward-looking.[6]

In the presence of a variance risk premium, the implied or risk-neutral variance is a biased estimate of the variance of the physical predictive density. Economic agents dislike the uncertainty of future variance and, in equilibrium, command a premium for accepting this risk, which gives rise to the variance risk premium. Bollerslev, Tauchen, and Zhou (2009) provide strong evidence of variance risk premia in financial assets. Thus, we first remove the variance risk premium from the risk-neutral variance before tilting the baseline predictive density $p(r_{t+1}|\mathcal{D}^t)$.

Let $\hat{u}_{\mathbb{P},t+1}$ denote the forecast error from the baseline physical predictive distribution at time $t+1$ obtained using either the SV or the GARCH model discussed in Section 1.2

$$\hat{u}_{\mathbb{P},t+1} = r_{t+1} - E(r_{t+1}|\mathcal{D}^t), \tag{20}$$

where $E(r_{t+1}|\mathcal{D}^t)$ is the posterior mean under $p(r_{t+1}|\mathcal{D}^t)$. The posterior variance of the predictive distribution is $\sigma_{\mathbb{P},t}^2 \equiv \text{Var}(r_{t+1}|\mathcal{D}^t)$. From options prices, we can compute the variance of the risk-neutral distribution, $\sigma_{\mathbb{Q},t}^2$, which differs from $\sigma_{\mathbb{P},t}^2$ by the variance risk premium VRP_{t+1}

$$\sigma_{\mathbb{P},t}^2 = \sigma_{\mathbb{Q},t}^2 - \text{VRP}_{t+1}. \tag{21}$$

---

6   Implied volatility reflects options traders' judgment about short-term volatility, due in part to information such as forthcoming announcements (e.g., an upcoming election, macroeconomic data releases) known to market participants but not to the econometrician. It resembles the judgmental component of Blue Chip, the Survey of Professional Forecasters, and the Greenbook surveys in forecasting INFL (among other macroeconomic series) as in Faust and Wright (2013).

We assume that the variance risk premium is such that the following holds

$$\log\left(\sigma_{\mathbb{P},t}^2\right) = \alpha + \beta \log\left(\sigma_{\mathbb{Q},t}^2\right). \tag{22}$$

Because the log-squared forecast error is a noisy measure of $\log \sigma_{\mathbb{P},t}^2$, we can estimate α and β using a regression of $\log\left(\hat{u}_{\mathbb{P},\tau+1}^2\right)$ on $\log\left(\sigma_{\mathbb{Q},\tau}^2\right)$, where $\tau = t_0, ..., t-1$. Thus, we tilt the predictive distribution such that its variance is given by

$$\hat{\sigma}_{\mathbb{P},t}^2 = \exp\left(\hat{\alpha} + \hat{\beta}\log\left(\sigma_{\mathbb{Q},t}^2\right)\right), \tag{23}$$

which implies that the variance risk premium is[7]

$$\begin{aligned}
\text{VRP}_{t+1} &= \sigma_{\mathbb{Q},t}^2 - \exp\left(\alpha + \beta\log\left(\sigma_{\mathbb{Q},t}^2\right)\right) \\
\hat{\text{VRP}}_{t+1} &= \sigma_{\mathbb{Q},t}^2 - \exp\left(\hat{\alpha} + \hat{\beta}\log\left(\sigma_{\mathbb{Q},t}^2\right)\right).
\end{aligned} \tag{24}$$

2 Empirical Results

We obtain the data necessary to generate the SV and GARCH density forecasts from Goyal and Welch (2008) and Rapach, Ringgenberg, and Zhou (2016). In what is by now a widely cited study, Goyal and Welch (2008) popularized a list of 14 predictors that capture fundamentals and have been used extensively in subsequent empirical asset pricing studies. Our end-of-month stock returns are computed from the S&P500 index and include dividends. We subtract a short T-bill rate from stock returns to obtain the monthly excess returns. Furthermore, we augment the set of the 14 popular predictor variables from Goyal and Welch (2008) with the short interest index (SII) introduced by Rapach et al.[8] Our sample starts in January 1973 ($t = 1$) and extends to December 2014 ($t = T$), as in Rapach et al.

We begin by computing the baseline predictive densities in Equation (17) and its GARCH analog for the equity premium using one by one, all 15 of the predictors considered. We also compute baseline predictive densities for the equity premium using a kitchen sink (KS) specification with 13 predictors (we exclude DE and TMS), as well as an equally weighted combination (EWC) of the predictive densities of all 15 predictors. To explicitly denote the dependence of the predictive density on model i, we write $p(r_{t+1}|\mathcal{M}_i, \mathcal{D}^t)$, where $i = 1, ..., K$ and $K = 17$. Next, to generate the predictive densities, we start in January 1986 and proceed in a recursive fashion using an expanding-window approach until the last

---

7 An alternative and more computationally demanding approach to incorporate forward-looking information into return forecasts would be to adapt an elaborate GARCH model like the MEM of Engle and Gallo (2006), the HEAVY of Shephard and Sheppard (2010), or the realized GARCH of Hansen, Huang, and Shek (2012). This adaptation would entail replacing realized volatility with a measure of implied volatility and developing an approach for handling the variance risk premium. See Table 1 in Hansen et al. for a succinct comparison of the three types of models.

8 The data on the monthly market returns, risk-free rate, and the Goyal and Welch (2008) predictors are available from Amit Goyal's website, updated and extended to December 2014, at http://www.hec.unil.ch/agoyal/ (accessed November 18, 2016). The SII data are available at http://sites.slu.edu/rapachde/home/research (accessed November 18, 2016). For a detailed discussion of the predictors considered, see Section 1 in Goyal and Welch (2008) and Section 2 in Rapach et al.
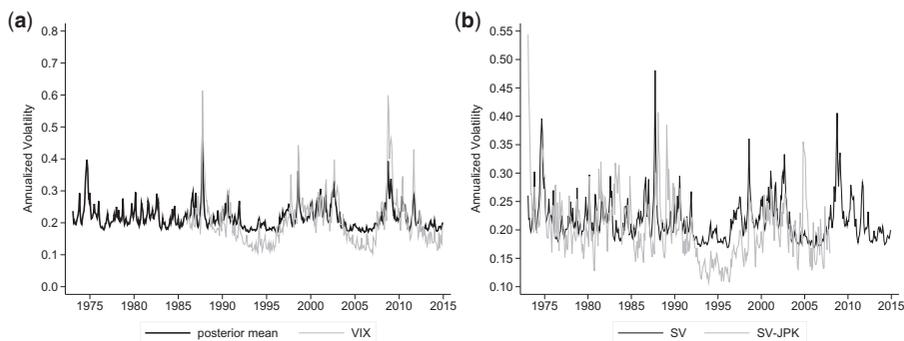
Figure 1. Annualized model-implied excess return volatility and VIX.

Notes: In panel (**a**), we show the annualized monthly excess return volatility implied by the SV model in Equations (8)–(9) for 1973:02–2014:12 using the SII as predictor. The black line shows the annualized posterior mean of $\exp(h_t)$, $t = 2, ..., T$. The gray line shows the annualized end-of-month values of the CBOE VIX for 1986:01–2004:12 using VXO for 1986:01–1989:12 as discussed in Section 2.2. In panel (**b**), we show the annualized monthly excess return volatility (SV) implied by the SV model in Equations (8) and (9) for 1973:02–2014:12 using the dividend yield (DY) as predictor, along with the SV series (SV–JPK) from Figure 4 in Johannes, Korteweg, and Polson (2014).

observation in the sample.[9] This process yields 17 time-series of one-step-ahead density forecasts—one for each predictor, one for the KS specification, and one for EWC—between January 1990 and December 2014.

In the interest of space, we will discuss the SV results more extensively than the GARCH results given that both models exhibit highly comparable statistical performance and the SV model performs better in terms of economic criteria. In Section C of the Online Appendix, we provide some more detailed results pertaining to the GARCH model.

2.1 Moments of the Physical and Risk-Neutral Distributions

To assess the degree of time variation in the excess return volatility implied by our econometric model, panel (a) of Figure 1 shows the monthly excess return volatility implied by the SV model in Equations (8)–(9) between February 1973 and December 2014. We plot the volatility series for a single predictor, the SII, noting that the series provided here are very similar across the predictors considered. The black line corresponds to the (full-sample) annualized posterior mean of $\exp(h_t)$. Although the annualized volatility hovers around 20% per month, it exhibits a couple of distinct spikes. The first one (40%) is in September 1974, which is 6 months after the end of the OPEC oil embargo in March 1974. The second corresponds to a value close to 48% in October 1987. The third, with a value of 39%, is in October 2008, amid the recent financial crisis.

As a comparison, we also plot the time-series of the end-of-month values of the Chicago Board Options Exchange (CBOE) Volatility Index (VIX). We use the VIX to summarize the risk-neutral volatility of the S&P500 returns, that is, the annualized $\sqrt{\sigma^2_{\mathbb{Q},t}}$. In 1993, the

---

9 Accordingly, we set aside the data from January 1973 to December 1985 to train the priors in Equations (10), (14), and (16). Hence, we set $t_0 = 156$.

CBOE introduced VIX, originally designed to measure the market's expectation of 30-day volatility implied by ATMS&P 100 Index (OEX) options prices. In 2003, CBOE together with Goldman Sachs updated the methodology and formula for VIX. The new VIX is based on the S&P500 Index (SPX) and estimates expected volatility by averaging the weighted prices of SPX puts and calls over a wide range of strikes and its values are available back to January 1990. We further extend the VIX series back to January 1986 by augmenting it with the VXO series. Setting aside the very prominent spikes in October 1987 and October 2008 that were also present in the series implied by our SV model, the risk-neutral volatility is highest during 1997–2003, a period of well-documented turmoil in financial markets (Bloom, 2009). Events during this period include the Asian crisis (Fall 1997), the Russian Financial Crisis (Fall 1998) September 11 (Fall 2001), the Enron and WorldCom scandals (Summer/Fall 2002), and Gulf War II (Spring 2003).

Based on a cursory look at panel (a) of Figure 1, the volatility series of our SV model lies above VIX during 1990–1997, as well as during 2004–2007, which are both periods of limited (if any) turmoil for the financial markets. This feature of the volatility series is not specific to our SV model. Panel (b) of Figure 1 makes this point by showing the volatility series from Figure 4 of Johannes, Korteweg, and Polson (2014). As can be inferred from this figure, their series is highly comparable in terms of magnitude and variation to the volatility series from our SV model.[10]

2.2 Entropic Tilting

We use the entropic tilting approach described in Section 1.1 to modify each of the baseline predictive densities, such that their variances match the corresponding option-implied risk-neutral variance—adjusted using Equation (23) to remove the variance risk premium—and their means are non-negative. Setting aside the period January 1986 to December 1989 to estimate the first variance risk premium, our final OOS period is January 1990 to December 2014.[11]

In Figure 2, we show the first two moments of the tilted and the baseline SV predictive densities over the OOS period for the model in which SII is the predictor. Starting with panel (a), the two mean series are essentially identical setting aside the differences due to numerical precision, except for the early part of the 1990s, between November 1990 and January 1991, and in August 2008, when we see a dip in the mean of the baseline return

---

10   We thank Johannes, Korteweg, and Polson (2014) for making their volatility series available to us. Although the volatility series from Johannes, Korteweg, and Polson (2014) spans the period December 1937–December 2007, here we plot it for February 1973 onward to be consistent with the beginning of our OOS period. We plot the volatility series from our SV model using the dividend yield as a predictor for two reasons. First, it makes essentially no difference for our volatility series shown in panel (a). Second, we want to be consistent with the predictor used by Johannes, Korteweg, and Polson (2014).

11   We obtain our estimates for the ith variance risk premium by regressing the log-squared forecast error implied by the ith baseline predictive density, $p(r_{\tau+1}|\mathcal{M}_i, \mathcal{D}^\tau)$, on the log-squared VIX using Equation (22) and an expanding-window approach, where $\tau = t_0, ..., t - 1$. Thus, the estimated variance risk premium for each forecast month comes from a regression using data from January 1986 through the previous month. For the SV model, the slope parameter $\hat{\beta}$ in Equation (22) has an average between 1.23 for DFY and 1.63 for the KS specification in these regressions. We refer the reader to Section B.1 of the Online Appendix for additional details.
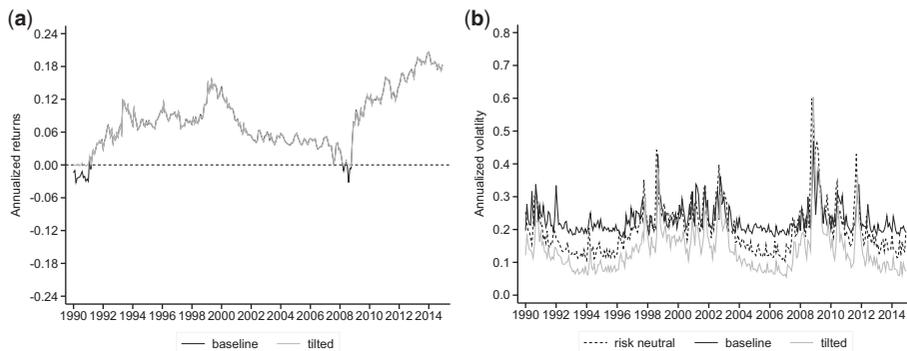
Figure 2. Moments of the baseline and tilted predictive densities, SV model.

Notes: The figure shows the first two moments of excess returns for the baseline and tilted predictive densities of the SV model in Equations (8) and (9) using the SII as predictor. Both predictive densities are OOS using recursive estimates for 1990:01–2014:12. Panel (**a**) shows the posterior mean of the baseline and tilted predictive densities. Panel (**b**) shows the posterior volatility of the baseline and tilted predictive densities along with end-of-month VIX values.

distribution but not in its tilted counterpart, an immediate consequence of the fact that we impose the non-negativity restriction on the mean of the tilted distribution. Panel (b) shows that the annualized volatility of the baseline distribution exceeds VIX for roughly 80% of the months between January 1990 and December 2014. Several months during 1997–2003 and 2008–2009 are exceptions. For example, the end-of-month value of VIX in October 2008 is close to 0.60, whereas the volatility of the baseline distribution is around 0.30. The volatility of the baseline distribution also exceeds its counterpart of the tilted distribution with a few exceptions, such as in the last 3 months of 2008 and in the early part of 2009. Importantly, the volatility of the baseline distribution in Figure 2 differs from the posterior mean of $\exp(h_t)$ shown in panel (a) of Figure 1 because it is based on recursive as opposed to full-sample estimates.

For all predictors, the shape of the baseline predictive densities is more dispersed over time compared with its tilted counterpart in the case of the SV model. Using SII as a predictor and if we focus on the far left tail of the distributions, the average 1% quantile for the baseline density forecasts is -0.259 while that for the tilted density forecasts is -0.097.[12] In the case of the far right tail of the distributions, the average 99% quantile for the baseline density forecasts is 0.273, while that for the tilted density forecasts is 0.111. Similar conclusions are drawn by looking at the shoulders of the two distributions. For example, the average 25% quantile of the baseline (tilted) density forecasts is -0.178 (-0.075). Similarly, the average 75% quantile of the baseline (tilted) is 0.043 (0.030).

The empirical KLIC defined in Equation (3) gauges how much the baseline density is altered by the tilting procedure. That is, small values of the empirical KLIC signify

---

12   Figures B-4 and B-5 of the Online Appendix to the paper contain fan charts for each of the predictors. Due to space limitations, we discuss here only the case of SII. The average value of the quantiles reported is calculated using 300 monthly observations between January 1990 and December 2014.

agreement between the baseline predictive model and outside information, while large values signify disagreement.[13] As a practical matter, large discrepancies also serve as warnings about the accuracy of statistics computed from the tilted densities. In fact, a large KLIC value implies that the distribution of the weights is highly skewed, with many draws from the baseline density being ignored and a few draws becoming highly influential.

For the SV model, the average KLIC values for the OOS period range from 0.095 to 0.110 depending on the predictor, which are comparable to the KLIC values reported in Cogley, Morozov, and Sargent (2005) and Robertson, Tallman, and Whiteman (2002), 0.12–0.68 and 0.06–0.66, respectively.[14] One of the three examples in Robertson et al. uses an intertemporal consumption-CAPM to add moment restrictions on a VAR forecasting real consumption growth and interest rates. This is the example that gives rise to the largest KLIC value reported in their paper (0.66), and—according to Cogley, Morozov, and Sargent (2005)—these values can serve as benchmarks for aggressive twisting given that the consumption-CAPM is known to fit the data poorly. In our case, although KLIC achieves some of its largest value during 1992–1995, 2004–2006, and 2012–2014, its annual average never exceeds 0.341 for any of the predictors in these years. Hence, it appears that for the largest part of our sample, the twisting of the baseline densities is not excessively aggressive.

3 OOS Performance

In this section, we examine whether the approach introduced in Section 1 leads to more accurate equity premium forecasts, both in terms of statistical and economic criteria. As with previous studies, such as Goyal and Welch (2008) and Campbell and Thompson (2008), we measure the predictive accuracy relative to appropriate HA models. However, since all the models we consider in this study allow for time-varying volatility, we augment the HA models to also include this feature and label them either HA–SV or HA–GARCH.

In particular, the HA–SV benchmark corresponds to the model in Equations (8) and (9) when $\beta = 0$.[15] The HA–GARCH benchmark corresponds to the model in Equations (18) and (19) when $\beta = 0$ using daily returns from which we subsequently simulate monthly returns. We construct the HA–GARCH benchmark using daily returns from which we subsequently simulate monthly returns. We do so because an HA–GARCH model estimated using daily returns provides a more stringent benchmark than the one estimated using monthly returns.[16]

---

13   Note that when $\text{KLIC}(\pi_t^*; \pi)$ is zero, it means that the baseline and tilted densities coincide.

14   The ranges reported here are based on Table 2 in Cogley, Morozov, and Sargent (2005), and on the KLIC statistics reported in Tables 1b, 2b, and 3b in Robertson et al.

15   For consistency, the HA–SV model is estimated using priors analogous to those we used with the various predictors. In particular, we slightly alter the prior on (μ, β) to impose a dogmatic "no predictability" prior on $\beta = 0$, while using the same priors for h^t, λ_0, λ_1, λ_2, and σ_ξ^{-2}.

16   For temporal aggregation of GARCH models as the one discussed here, see Zivot (2009) and Heston and Nandi (2000), among others. Note that we will also compare the SV model against the HA–GARCH benchmark in a later section.

3.1 Statistical Forecasting Performance

We consider several evaluation metrics for both point and density forecasts. Starting with point-forecast accuracy, we follow Campbell and Thompson (2008) and summarize the predictive ability of the various models over the whole evaluation period by reporting the OOS R-squared for the forecasting model associated with each model k

$$R^2_{\text{OOS},kd} = 1 - \frac{\sum_{\tau=m+1}^{T} e^2_{kd,\tau}}{\sum_{\tau=m+1}^{T} e^2_{bcmk,\tau}}, \tag{25}$$

where $m+1$ denotes the beginning of the forecast evaluation period (January 1990) and $bcmk$ refers to the appropriate (HA–SV or HA–GARCH) benchmark. The additional subscript $d \in \{\text{baseline, tilted}\}$ allows us to distinguish between the baseline and the tilted densities. Furthermore, $e_{kd,\tau}$ and $e_{bcmk,\tau}$ denote the time τ forecast error for the baseline or tilted, and the corresponding benchmark densities, respectively. We obtain point forecasts to compute the forecast errors in Equation (25) by averaging over the draws from the corresponding predictive densities. A positive $R^2_{\text{OOS},kd}$ indicates that the point forecasts associated with the baseline or tilted densities are, on average, more accurate than the benchmark forecasts.

To quantify the accuracy of density forecasts, we follow Amisano and Giacomini (2007) and report the average log score difference (ALSD)

$$\text{ALSD}_{kd} = \frac{1}{T-m} \sum_{\tau=m+1}^{T} \text{LS}_{kd,\tau} - \text{LS}_{bcmk,\tau}, \tag{26}$$

where $\text{LS}_{kd,\tau}$ and $\text{LS}_{bcmk,\tau}$ denote the time-τ log predictive scores of the baseline or tilted densities, and the predictive density, respectively. The logarithmic score gives a high value to a predictive density that assigns a high probability to the event that actually occurred. Hence, a positive ALSD_{kd} value indicates that, on average, the SV (GARCH) model is more accurate than the HA–SV (HA–GARCH) benchmark in predicting the outcome of interest.[17]

To test the statistical significance of differences in point and density forecasts, we consider Diebold and Mariano (1995) (DM) tests of equal predictive accuracy using mean-squared forecast errors (MSFEs) and average log scores (ALSs), respectively. We perform two DM tests. First, we test whether the improvements in the MSFEs or the ALSs for the baseline densities relative to their HA–SV benchmark counterparts are statistically significant. Second, we test whether the improvements in the MSFEs or the ALSs for the tilted densities relative to their baseline counterparts are statistically significant. In both cases, we use standard normal critical values and incorporate the finite sample correction due to Harvey, Leybourne, and Newbold (1997).[18]

The top panel of Table 1 pertains to point forecasts from the SV model. Columns (1) and (2) of the table report the R^2_{OOS} (in percent) associated with the baseline and tilted

---

17 We compute the log predictive score by relying on a kernel-smoothing technique to estimate the predictive density at its realized values.

18 Citing Monte-Carlo evidence in Clark and McCracken (2011), with nested models, Clark and Ravazzolo (2015) argue that the DM test with normal critical values is a somewhat conservative test—has sizes that tend to fall below the nominal—for equal accuracy in finite samples.

Table 1. OOS statistical predictability, SV model

Panel A: OOS R^2 (vs. HA–SV)

| | 1990:01–2014:12 | | 1990:01–2006:12 | | 2007:01–2014:12 | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
Model	Baseline	Tilted	Baseline	Tilted	Baseline	Tilted
DP	−1.153	**−1.074**	−1.876	**−1.750**	−0.008	**−0.004**
DY	−1.189	**−1.044**	−2.026	**−1.829**	0.138	**0.200**
EP	−0.654	**−0.464**	−0.879	**−0.863**	−0.297	**0.167**
DE	−1.579	**−1.040**	−1.816	**−1.774**	−1.204	**0.123**
RVOL	−0.711	**−0.295***	−2.003	**−1.361**	1.335	**1.392**
BM	−0.688	−0.711	−1.010	−1.033	−0.178	−0.200
NTIS	−3.070	**−2.321****	−2.305	**−1.119****	−4.281	**−4.223**
TBL	−0.761	−0.830	−0.978	−1.011	−0.418	−0.543
LTY	−0.320	−0.326	−0.578	−0.579	0.090	0.076
LTR	−0.270	−0.359	−0.466	−0.500	0.039	−0.134
TMS	−0.703	−0.734	−0.901	**−0.886**	−0.390	−0.493
DFY	−1.370	**−0.957***	−1.328	**−0.738***	−1.437	**−1.305**
DFR	−0.813	−1.543	−2.018	**−1.275**	1.096	−1.966
INFL	−1.081	**−1.034**	−0.371	**−0.271**	−2.205	−2.243
SII	1.497*	**1.524**	0.136	**0.195**	3.653*	3.629
KS	−13.717	**−4.829**	−7.015	**−2.738****	−24.330	**−8.141**
EWC	−0.064	−0.064	−0.309	−0.309	0.324	0.324

Panel B: Average log score differences (vs. HA–SV)

| | 1990:01–2014:12 | | 1990:01–2006:12 | | 2007:01–2014:12 | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
Model	Baseline	Tilted	Baseline	Tilted	Baseline	Tilted
DP	−0.008	**0.200*****	−0.011	**0.201*****	0.000	**0.199*****
DY	−0.009	**0.199*****	−0.014	**0.199*****	0.000	**0.200*****
EP	−0.003	**0.207*****	−0.005	**0.210*****	0.002	**0.200*****
DE	−0.011	**0.189*****	−0.015	**0.184*****	−0.002	**0.199*****
RVOL	−0.010	**0.201*****	−0.020	**0.198*****	0.010	**0.209*****
BM	0.000	**0.205*****	−0.003	**0.208*****	0.005**	**0.199*****
NTIS	−0.011	**0.195*****	−0.014	**0.205*****	−0.004	**0.173*****
TBL	0.003	**0.199*****	−0.003	**0.198*****	0.015**	**0.203*****
LTY	0.001	**0.206*****	−0.003	**0.209*****	0.009**	**0.201*****
LTR	−0.002	**0.209*****	−0.003	**0.213*****	0.001	**0.200*****
TMS	−0.003	**0.190*****	−0.009	**0.184*****	0.009*	**0.204*****
DFY	−0.010	**0.202*****	−0.014	**0.204*****	−0.001	**0.198*****
DFR	−0.005	**0.193*****	−0.006	**0.197*****	−0.004	**0.184*****
INFL	0.001	**0.209*****	−0.002	**0.212*****	0.007**	**0.203*****
SII	0.005	**0.216*****	−0.003	**0.216*****	0.022**	**0.216*****
KS	−0.029	−1.708	−0.034	−2.581	−0.019	**0.149****
EWC	−0.003	**0.211*****	−0.006	**0.214*****	0.005**	**0.204*****

Notes: The table reports the OOS R^2 in Equation (25) and the ALSDs in Equation (26) for the 17 models considered, over the entire OOS period, 1990:01–2014:12, as well as for 1990:01–2006:12 and 2007:01–2012:14. All forecasts are OOS using recursive estimates for 1990:01–2014:12. Bold numbers indicate all instances where the tilted forecasts improve upon the corresponding baseline forecasts. *,**,***Statistical significance at 10, 5 and 1% levels, using the DM (1995) tests discussed in Section 3.1. The model nomenclature is as follows: (1) DP, log dividend price-ratio; (2) DY, log dividend yield; (3) EP, log EP ratio; (4) DE, log dividend-payout ratio; (5) RVOL, excess stock return volatility; (6) BM, book-to-market ratio; (7) NTIS, net equity expansion; (8) TBL, treasury bill rate; (9) LTY, long-term yield; (10) LTR, long-term return; (11) TMS, term spread; (12) DFY, default yield spread; (13) DFR, default return spread; (14) INFL, inflation; (15) SII, short interest index; (16) KS, kitchen sink; (17) EWC, equally weighted combination. We use HA–SV to refer to the historical-average model augmented with SV.

density forecasts for each of the predictors, as well as the KS specification and the equally weighted forecast combination (EWC), over the full OOS period, January 1990–December 2014. The remaining columns report the R^2_{OOS} values for the earlier (January 1990–December 2006) and later (January 2007–December 2014) parts of the OOS period. For example, SII produces an R^2_{OOS} of 1.497% in the case of the baseline forecasts and an R^2_{OOS} of 1.524% in the case of the tilted forecasts for the full OOS period. The bold entry in Column (2) indicates that the tilted forecasts perform better than the baseline forecasts in terms of MSFEs generating a higher R^2_{OOS}. The lack of an asterisk next to the entry in Column (2) for the same predictor indicates that the tilted forecasts fail to be significantly better than the baseline forecasts. Analogous notational conventions hold for the other combinations of models and OOS periods in the table.

In the case of baseline point forecasts for the full OOS period, we observe negative R^2_{OOS} values for all models except for SII. Consistent with overfitting, the KS specification is rather disappointing delivering a R^2_{OOS} of −13.717%. These results are consistent with the findings of Rapach, Ringgenberg, and Zhou (2016). Although tilting leads to R^2_{OOS} improvements for 10 out of 17 models (1 of them is significant at 5%), it fails to produce positive R^2_{OOS} values with the exception of SII.

The bottom panel of Table 1 reports the ALSDs for the baseline and tilted density forecasts. Over the full OOS period, T-bill (TBL), long-term yield (LTY), inflation (INFL), and SII are the only predictors for which the ALSD associated with the baseline density forecasts is positive. The remaining ALSDs lie between −0.029 for the KS specification and zero for the book-to-market (BM) ratio. The tilting procedure delivers a substantial improvement in the ALSDs for all models except for the KS specification. The resulting improvements are all statistically significant at the 1% level. Excluding the KS specification, we see ALSD values associated with the tilted density forecasts between 0.189 for the dividend-payout ratio (DE) and 0.216 for SII.

The four leftmost columns of Table 3 pertain to the statistical performance of the GARCH model. Starting with the R^2_{OOS} metrics, we see that the baseline point forecasts improve over the HA–GARCH benchmark specification in 9 out of the 17 models. These improvements, however, are not statistically significant. On the other hand, tilting entails improvements in R^2_{OOS} for 12 of the 17 models. Once again, the differences between the tilted and baseline point forecasts are not statistically significant. As for the accuracy of the density forecasts, we first notice that the baseline ALSDs are all negative, indicating that the GARCH baseline models fail to produce more accurate density forecasts than those implied by the benchmark HA–GARCH model. On the other hand, it appears that our tilting procedure improves the ALSDs for all 17 models and, in four cases, in a statistically significant way (5%). Finally, for 13 of these 17 cases, tilting leads to density forecasts that are more accurate than those implied by the benchmark HA–GARCH model.

To see how point-forecast performance changes over time, we compute the cumulative sum of squared forecast error difference (CSSED)

$$\mathrm{CSSED}_{kd,t} = \sum_{\tau=m+1}^{t} \left(\mathrm{e}^2_{bcmk,\tau} - \mathrm{e}^2_{kd,\tau} \right), \quad t = m+1, ..., T. \tag{27}$$
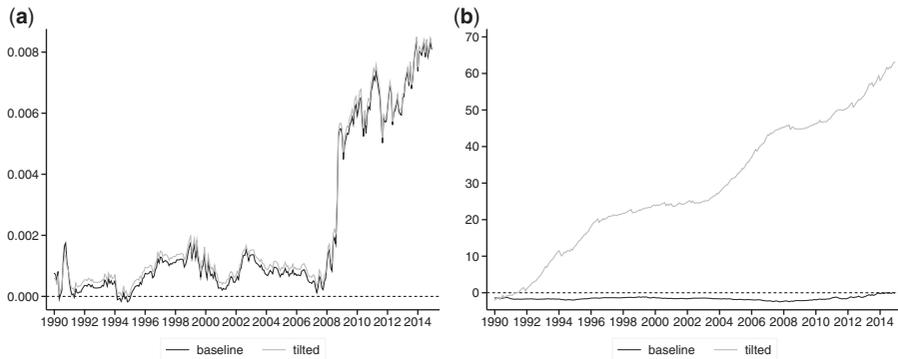
Figure 3. Cumulative sums of squared forecast error and log score differences, SV model.

Notes: Panel (**a**) shows the cumulative sum of squared error differences (CSSEDs) in Equation (27) using the SII as a predictor. CSSEDs above zero indicate that the baseline and/or tilted densities generate better predictions than the historical-average with SV (HA–SV) benchmark densities. Negative CSSED values suggest the opposite. Panel (**b**) shows the cumulative log score differences (CLSDs) in Equation (28) using the same predictor. CLSDs above zero indicate that the baseline and/or tilted models generate better performance than the HA–SV benchmark. Negative values suggest the opposite. All forecasts are OOS using recursive estimates for 1990:01–2014:12.

A positive $\text{CSSED}_{kd,t}$ indicates that up to time t the point forecasts associated with the baseline or tilted predictive densities for model k are more accurate, on average, than their benchmark (HA–SV or HA–GARCH) counterparts. We also examine how the density-forecast performance changes over time using the cumulative log-score difference (CLSD)

$$\text{CLSD}_{kd,t} = \sum_{\tau=m+1}^{t} \text{LS}_{kd,\tau} - \text{LS}_{bcmk,\tau}, \quad t = m+1, ..., T. \tag{28}$$

If the baseline or tilted density forecasts are more accurate than the benchmark ones throughout the entire OOS period, then the corresponding CLSD line would be monotonically increasing. Conversely, episodes with the density forecasts being less accurate than the benchmark would generate dips in the CLSD line.

Panel (a) of Figure 3 plots the CSSED associated with the baseline and tilted SV density forecasts for SII and illustrates the role of the non-negativity constraint in our tilting procedure. For SII, as panel (a) of Figure 2 shows, the point forecasts turn negative for a short period of time; in the early 1990s and around the latest financial crisis. The fact that point forecasts turn negative in the early 1990s creates an initial wedge between the CSSED series that is maintained until the end of the OOS period. Around 2008, the onset of the latest financial crisis leads to a large positive shock in predictability for both the baseline and tilted models. Although the CSSED for the tilted density forecasts wins the horse race, its terminal value is not significantly larger than that of its counterpart for the baseline density forecasts. Panel (b) of Figure 3 plots the CLSDs for the SII baseline and tilted densities. The tilted CLSD line lies above the baseline one throughout the OOS period with a clear upward trend that leads to a terminal value in excess of 60.

The baseline CLSD line, on the other hand, remains very close to zero throughout the entire OOS period.

We conclude this section by investigating the stability over time of the point and density forecasts for the SV model.[19] In particular, we perform two separate analyses. First, we separately report the R^2_{OOS} and ALSD statistics for two different parts of the OOS period. The first part, January 1990–December 2006 (OOS-I), predates the global financial crisis. The second part, January 2007–December 2014 (OOS-II), surrounds the recent crisis. We also report the results of the Giacomini and Rossi (2010) fluctuation test for the baseline and tilted density forecasts in terms of MSFE differences (MSFEDs) and ALSDs.

We begin with the results in Columns (3)–(6) of Table 1. Similar to the full OOS results, tilting leads to improvements in R^2_{OOS} for both OOS-I and OOS-II periods. In the case of period OOS-I, the tilted R^2_{OOS} values are higher than their benchmark counterparts for 12 out of the 17 models; the improvements for two models are statistically significant at 5%. Turning to period OOS-II, we notice that the baseline point forecasts imply a positive R^2_{OOS} value in several instances, ranging between 0.039 for the long-term return (LTR) and 3.653 for SII. The improvement due to altering the first moment of the baseline densities is limited to eight models. Moving to panel (b) of Table 1, we find that the effect of tilting on the density forecast accuracy is beneficial for both periods. In every instance, the ALSD values for the tilted densities are higher than their baseline counterparts. Furthermore, in all instances the differences are statistically significant at the 1% level, except for the KS specification.

Next, Figure 4 plots the results of the Giacomini–Rossi (GR) fluctuation test for the baseline and tilted densities, both in terms of MSFEDs and ALSDs. For the baseline densities, we test the null hypothesis that the baseline and the HA–SV densities have equal predictive performance at each point in time over 5-year centered windows in our OOS period. The alternative hypothesis is that the baseline densities perform better. For the tilted densities, we test the null hypothesis that the tilted and baseline densities have equal predictive performance. The alternative hypothesis is that the tilted densities perform better. We test these hypotheses for each of the 17 cases we consider. As a result, the maximum number of rejections reported is 17.

If the forecasting performance is stable over time, we expect the rejection rate to be relatively constant over time. Starting with panels (a) and (c), which use the MSFED metric, we see that for the baseline predictive densities we reject the null hypothesis only for a few cases during 1998–2000 and 2009–2011. For the same metric, we see mostly between 1 and 6 rejections for the tilted densities with the rejections clustering primarily during 1998–2008.

Using the ALSD metric, we generally fail to reject the null hypothesis for the baseline densities except for the periods 1998–2000 and 2012–2014 (panel (b)). During the earlier period, the number of rejections is fairly small and tends to not exceed 5. During the later period, the number of rejections is generally larger and lies somewhere between 2 and 8. For the tilted densities, we consistently reject the null for almost all models with the exception of a short window during 2002–2004, where we reject the null for no more than 7 models [panel (d)]. In sum, the improvement in logarithmic scores due to tilting is

---

19 We provide a similar breakdown for GARCH in Section C of the Online Appendix.
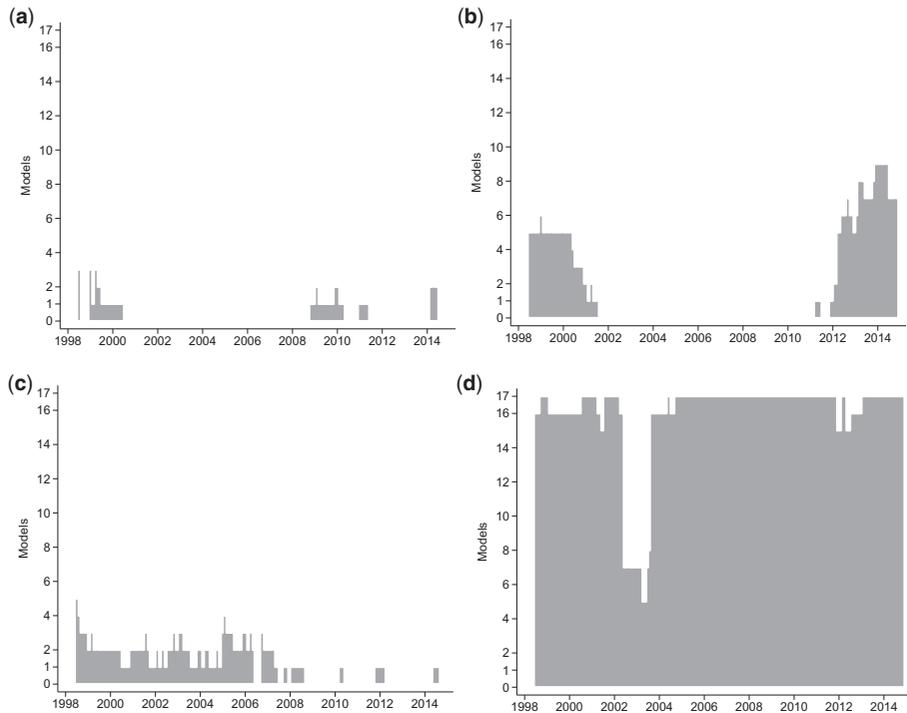
Figure 4. Fluctuation test across predictors, SV model.

Notes: The figure shows the number of rejections in the one-sided Giacomini and Rossi (2010) fluctuation test for the baseline and tilted predictive densities of all 17 models considered over centered rolling windows of 60 observations as discussed in Section 3.1. The critical values have been adjusted to reflect our recursive forecasts. The two left panels focus on the mean-squared forecast error differences (MSFEDs), while the two right panels focus on the average log-score differences (ALSDs) in Equation (26). All forecasts are OOS using recursive estimates for 1990:01–2014:12.

consistently strong for the majority of the models considered and for almost the entirety of the OOS period and does not appear to be driven by a few isolated events.

3.2 Economic Performance

Up to this point, we have focused on the statistical performance of the baseline and tilted predictive densities. In this section, we turn to their economic performance. We posit a representative investor using these predictions to make optimal portfolio decisions while taking parameter uncertainty into consideration.[20]

In particular, our interest lies in the optimal asset allocation of a representative investor facing a utility function $U(\omega_{t-1}, r_t)$ with ω_{t-1} denoting the share of her portfolio allocated

---

20 Our discussion follows closely Kandel and Stambaugh (1996) and Barberis (2000). Parameter uncertainty is accounted for in the Bayesian framework because the parameter posterior distribution is integrated out of the predictive density of returns [see Equation (17)].

into risky assets, and r_t being the equity premium at time t.[21] The representative agent solves the optimal asset allocation problem

$$\omega_{t-1}^* = \arg \max_{\omega_{t-1}} E[U(\omega_{t-1}, r_t)|\mathcal{D}^{t-1}], \tag{29}$$

with $t = m+1, ..., T$. She is assumed to have power utility of the form

$$U(\omega_{t-1}, r_t) = \frac{\left[(1-\omega_{t-1})\exp\left(r_{f,t-1}\right) + \omega_{t-1}\exp\left(r_{f,t-1} + r_t\right)\right]^{1-A}}{1-A}, \tag{30}$$

where $r_{f,t-1}$ is the continuously compounded T-bill rate available at time $t-1$, and A is the coefficient of relative risk aversion. The subscript $t-1$ on the portfolio implies that the investor solves the optimization problem using information available only at time $t-1$. The power utility function exhibits the useful property of constant relative risk aversion (CRRA). Moreover, the optimal portfolio weights do not depend on initial wealth.

Taking expectations with respect to the predictive density of r_t, we can rewrite Equation (29) as follows:

$$\omega_{t-1}^* = \arg \max_{\omega_{t-1}} \int U(\omega_{t-1}, r_t)p(r_t|\mathcal{D}^{t-1})dr_t. \tag{31}$$

The integral in Equation (31) can be approximated using draws from the competing predictive densities. Specifically, using the HA–SV or HA–GARCH predictive density, we can approximate the solution to Equation (31) with a large number (J) of draws, $\{r_{bcmk,t}^j\}_{j=1}^J$, and the following expression[22]:

$$\hat{\omega}_{bcmk,t-1} = \arg \max_{\omega_{t-1}} \frac{1}{J} \sum_{j=1}^{J} \left\{ \frac{\left[(1-\omega_{t-1})\exp\left(r_{f,t-1}\right) + \omega_{t-1}\exp\left(r_{f,t-1} + r_{bcmk,t}^j\right)\right]^{1-A}}{1-A} \right\}. \tag{32}$$

Similarly, using kd with $d \in \{\text{baseline, tilted}\}$ to denote either the baseline or the tilted density forecasts for model k, we can approximate Equation (31) via

$$\hat{\omega}_{kd,t-1} = \arg \max_{\omega_{t-1}} \frac{1}{J} \sum_{j=1}^{J} \left\{ \frac{\left[(1-\omega_{t-1})\exp\left(r_{f,t-1}\right) + \omega_{t-1}\exp\left(r_{f,t-1} + r_{kd,t}^j\right)\right]^{1-A}}{1-A} \right\}. \tag{33}$$

The sequence of portfolio weights $\{\hat{\omega}_{bcmk,t-1}\}_{t=m+1}^T$ and $\{\hat{\omega}_{kd,t-1}\}_{t=m+1}^T$ is next used to compute the realized utilities for the benchmark, baseline, and tilted densities. Let $\hat{W}_{bcmk,t}$ and $\hat{W}_{kd,t}$ be the corresponding realized wealth at time t, where $\hat{W}_{bcmk,t}$ and $\hat{W}_{kd,t}$ are functions

21 Given the availability of density forecasts as opposed to just point forecasts, we are not restricted to rely on a mean–variance utility function, and we can focus on functions with better properties such as the power utility. The power utility avoids the major limitation of the mean–variance utility, namely, that investors care only about the first two moments of returns. Furthermore, it is well known that mean–variance portfolio optimization is consistent with expected utility maximization only under special circumstances. Sufficient conditions include quadratic utility or elliptical return distributions. See, for example, Back (2010).

22 As described in footnote 4, we set $J = 5000$.

of time t realized excess return, r_t, as well as the optimal allocations to the risky and risk-free assets computed in Equations (32) and (33)

$$
\begin{aligned}
\hat{W}_{bcmk,t} &= (1 - \hat{\omega}_{bcmk,t-1})\exp\left(r_{f,t-1}\right) + \hat{\omega}_{bcmk,t-1}\exp\left(r_{f,t-1} + r_t\right) \\
\hat{W}_{kd,t} &= (1 - \hat{\omega}_{kd,t-1})\exp\left(r_{f,t-1}\right) + \hat{\omega}_{kd,t-1}\exp\left(r_{f,t-1} + r_t\right).
\end{aligned}
\tag{34}
$$

Following Cenesizoglu and Timmermann (2012), we assess the performance of the predictive densities by calculating the implied annualized certainty equivalent return (CER) values for the OOS period as follows:

$$
\begin{aligned}
\text{CER}_{bcmk} &= \left((1 - A)(T - m)^{-1}\sum_{\tau=m+1}^{T}\hat{U}_{bcmk,\tau}\right)^{12/(1-A)} - 1 \\
\text{CER}_{kd} &= \left((1 - A)(T - m)^{-1}\sum_{\tau=m+1}^{T}\hat{U}_{kd,\tau}\right)^{12/(1-A)} - 1,
\end{aligned}
\tag{35}
$$

where $\hat{U}_{bcmk,\tau} = \hat{W}_{bcmk,\tau}^{1-A}/(1 - A)$ and $\hat{U}_{kd,\tau} = \hat{W}_{kd,\tau}^{1-A}/(1 - A)$ denote the time-τ realized utility associated with the benchmark and the baseline or tilted predictive density, respectively. Finally, we compute the certainty equivalent return difference (CERD) using

$$
\text{CERD}_{kd} = \text{CER}_{kd} - \text{CER}_{bcmk}.
\tag{36}
$$

Table 2 reports the annualized CERD estimates associated with the baseline and tilted SV density forecasts. For the remainder of our discussion here, we will refer to the former as baseline CERDs and we will refer to the latter as tilted CERDs. As in Table 1, we separately report results for the entire OOS period as well as for the two shorter periods, January 1990–December 2006 (OOS-I) and January 2007–December 2014 (OOS-II). We also examine the sensitivity of the CERDs to different risk preferences by considering risk-aversion coefficients of 3 (top panel) and 5 (bottom panel).[23]

Starting with $A = 3$ and the full OOS period, the tilted CERDs exceed the corresponding baseline CERDs for all 17 models considered [panel (a)]. Across the 17 models, the average baseline CERD is 0.043% with the model-specific CERDs ranging between −0.849% for net equity expansion (NTIS) and 1.310% for SII. The average tilted CERD is 2.279% with model-specific values between 1.151% for default yield spread (DFY) and 4.890% for SII. We see an average increase of 224 basis points (bps) relative to the baseline CERDs calculated using the difference between Columns (2) and (1). For the OOS-I period, the tilted CERDs exceed the baseline ones in all but one model, with an average improvement over the baseline CERDs of 150 bps. The largest improvements relative to the baseline CERDs, 4.112%, are associated with the TMS. For the OOS II period, the tilted CERDs exceed the baseline CERDs for all 17 models. The average increase relative to the baseline CERDs is almost 380 bps, with the tilted density forecasts giving rise to improvements as high as

---

23 We compute the optimal portfolio weights for the CRRA investor using the approximation in Equation 2.4 of Campbell and Viceira (2002). Additionally, we restrict the portfolio weights to lie between −0.5 and 1.5 as in Rapach, Ringgenberg, and Zhou (2016). We have also experimented with tighter bounds on the portfolio weights, ruling out short-selling and leverage (i.e., $\omega_t \in [0, 1)$), as well as fully unconstrained portfolio weights. The results from these experiments are qualitatively very similar to the main results we report in Table 2.

Table 2. OOS economic predictability, SV model

Panel A: CER gains, $A = 3$ (vs. HA–SV)

	1990:01–2014:12		1990:01–2006:12		2007:01–2014:12	
	(1)	(2)	(3)	(4)	(5)	(6)
Model	Baseline	Tilted	Baseline	Tilted	Baseline	Tilted
DP	−0.514	**1.344**	−0.694	**0.201**	−0.141	**3.743**
DY	−0.510	**1.459**	−0.718	**0.329**	−0.081	**3.832**
EP	−0.155	**1.830**	−0.213	**0.909**	−0.036	**3.756**
DE	−0.061	**2.618**	−0.002	**2.076**	−0.181	**3.745**
RVOL	−0.098	**1.317**	−0.600	−0.843	0.944	**5.933**
BM	0.111	**1.727**	0.145	**0.934**	0.040	**3.381**
NTIS	−0.849	**1.994**	−0.640	**1.036**	−1.279	**4.000**
TBL	0.750	**2.464**	0.576	**1.856**	1.110	**3.729**
LTY	0.426	**1.835**	0.245	**1.105**	0.799	**3.356**
LTR	−0.099	**1.492**	−0.014	**0.788**	−0.274	**2.959**
TMS	0.216	**4.100**	0.184	**4.296**	0.280	**3.695**
DFY	−0.747	**1.151**	−0.827	−0.051	−0.581	**3.679**
DFR	−0.104	**3.685**	−0.720	**3.180**	1.176	**4.732**
INFL	0.129	**1.984**	0.335	**1.305**	−0.293	**3.397**
SII	1.405	**4.890**	0.251	**2.954**	3.829	**9.010**
KS	0.726	**2.144**	−0.331	**0.203**	2.942	**6.276**
EWC	0.106	**2.709**	−0.047	**2.154**	0.423	**3.862**

Panel B: CER gains, $A = 5$ (vs. HA–SV)

	1990:01–2014:12		1990:01–2006:12		2007:01–2014:12	
	(1)	(2)	(3)	(4)	(5)	(6)
Model	Baseline	Tilted	Baseline	Tilted	Baseline	Tilted
DP	−0.319	**0.710**	−0.412	−0.557	−0.129	**3.366**
DY	−0.302	**0.749**	−0.437	−0.569	−0.024	**3.514**
EP	−0.082	**2.058**	−0.117	**1.209**	−0.011	**3.824**
DE	−0.033	**1.583**	0.005	**0.764**	−0.109	**3.284**
RVOL	−0.060	**0.770**	−0.357	−1.236	0.550	**5.037**
BM	0.086	**1.496**	0.109	**0.829**	0.041	**2.877**
NTIS	−0.494	**0.938**	−0.366	**0.832**	−0.754	**1.154**
TBL	0.453	**1.831**	0.327	**1.511**	0.711	**2.488**
LTY	0.284	**1.765**	0.174	**1.146**	0.511	**3.046**
LTR	−0.037	**1.482**	0.002	**1.166**	−0.114	**2.129**
TMS	0.161	**2.512**	0.134	**2.510**	0.214	**2.516**
DFY	−0.432	**0.621**	−0.483	−0.204	−0.329	**2.335**
DFR	−0.018	**2.934**	−0.408	**1.747**	0.783	**5.417**
INFL	0.091	**1.987**	0.217	**1.356**	−0.168	**3.293**
SII	0.849	**4.199**	0.159	**2.359**	2.278	**8.100**
KS	0.252	**1.448**	−0.200	−0.315	1.182	**5.179**
EWC	0.066	**2.457**	−0.020	**1.665**	0.244	**4.100**

Notes: The table reports the annualized CERDs in Equation (36) for portfolio decisions based on recursive OOS forecasts of excess returns. Each period, an investor with power utility and coefficient of relative risk aversion $A = 3$ (top panel) or $A = 5$ (bottom panel) selects stocks and T-bills based on a predictive density differing both by the model considered and the predictive density entertained (baseline or tilted). See the notes of Table 1 for the model nomenclature. The equity weights are constrained to lie in the $[−0.5, 1.5]$ interval. All forecasts are OOS using recursive estimates of the models for 1990:01–2014:12. Bold numbers indicate all instances where CER gains for the tilted densities exceed the CER gains for the baseline densities. We use HA–SV to refer to the historical-average model augmented with SV.

Table 3. OOS statistical and economic predictability, GARCH(1,1) model

Model	OOS R^2		ALSD		CER gains, $A = 3$		CER gains, $A = 5$	
	(1) Baseline	(2) Tilted	(3) Baseline	(4) Tilted	(5) Baseline	(6) Tilted	(7) Baseline	(8) Tilted
DP	−0.426	**0.142**	−0.093	**−0.002**	−1.106	**0.563**	−0.827	**0.386**
DY	−0.484	**0.236**	−0.049	**0.042**	−0.894	**0.743**	−0.659	**0.500**
EP	0.355	**1.031**	−0.040	**0.052***	1.757	**4.057**	1.041	**3.011**
DE	−0.629	**0.334**	−0.049	**0.018**	2.113	**3.972**	1.116	**2.224**
RVOL	−0.004	**0.335**	−0.061	**0.058***	0.024	**1.810**	0.017	**0.905**
BM	0.729	**0.735**	−0.041	**0.046**	1.435	**2.824**	0.589	**2.257**
NTIS	−1.168	**−0.327**	−0.081	**0.001**	−0.200	**1.032**	−0.871	**0.179**
TBL	0.898	**0.906**	−0.070	**−0.023**	1.853	**2.535**	1.412	**2.317**
LTY	1.023	1.020	−0.046	**−0.001**	1.968	**2.500**	1.151	**2.363**
LTR	0.106	**0.291**	−0.042	**0.016**	0.181	**0.994**	0.365	**1.221**
TMS	0.248	0.207	−0.051	**0.016**	1.182	**2.880**	0.644	**2.014**
DFY	−1.120	**−0.312**	−0.072	**0.058***	−2.128	**1.184**	−1.328	**0.671**
DFR	−0.081	−0.762	−0.037	**0.068***	2.254	**4.560**	0.632	**2.975**
INFL	0.453	**0.566**	−0.036	**0.052**	1.468	**1.632**	0.945	**1.783**
SII	1.899	1.790	−0.044	**0.032**	3.456	**4.693**	2.706	**3.515**
KS	−22.973	**−14.158**	−0.188	**−0.018**	−0.618	**0.659**	−1.595	**0.623**
EWC	0.954	0.954	−0.034	**0.075***	1.962	**3.627**	1.294	**2.649**

Notes: The table shows results for the OOS statistical and economic predictability of the GARCH(1,1) model using the HA–GARCH benchmark for 1990:01–2014:12. Bold numbers indicate all instances where the tilted forecasts improve upon the corresponding baseline forecasts. *,**,***Statistical significance at 10, 5 and and 1% levels respectively, using the DM (1995) tests discussed in Section 3.1.

5.279% relative to their baseline counterparts in the case of SII. In the case of $A = 5$ in panel (b), the average improvement relative to the baseline CERDs one would obtain by tilting is 171 bps for 1990–2014, 93.4 bps for the OOS-I period, and 334 bps for the OOS-II period.

As for the economic predictability of the GARCH model, it appears that in all cases (and across different risk aversion choices) tilting the baseline GARCH density forecasts leads to large CER gains; refer to Columns (5)–(8) of Table 3. In the case of $A = 3$, the baseline CERDs range between −2.128% for the default spread yield (DFY) and 3.456% for SII with an average of 0.865%. Their tilted analogs are 0.563% for the dividend-price ratio (DP) and 4.693% for SII with an average of 2.369%. Hence, tilting leads to an improvement of close to 150 bps. When we consider the case of $A = 5$, the baseline CERDs are between −1.595% (KS) and 2.706 (SII) with an average of 0.390%. The tilted CERDs are now between 0.179% for the net equity expansion (NTIS) and 3.515% for SII with an average of 1.741%. Hence, tilting leads to an improvement of about 135 bps.

Finally, the top 2 panels of Figure 5 plot the time-series of equity allocation weights for the monthly portfolios based on the EP and SII baseline and tilted SV densities, along with the equity weights implied by the HA–SV benchmark densities, assuming $A = 3$. While the HA–SV equity weights oscillate between 0.230 and 0.560, with an average of 0.398, the baseline and tilted equity weights exhibit more variation. This is especially true for the tilted equity weights between 1998 and 2003 and right after the financial crisis.
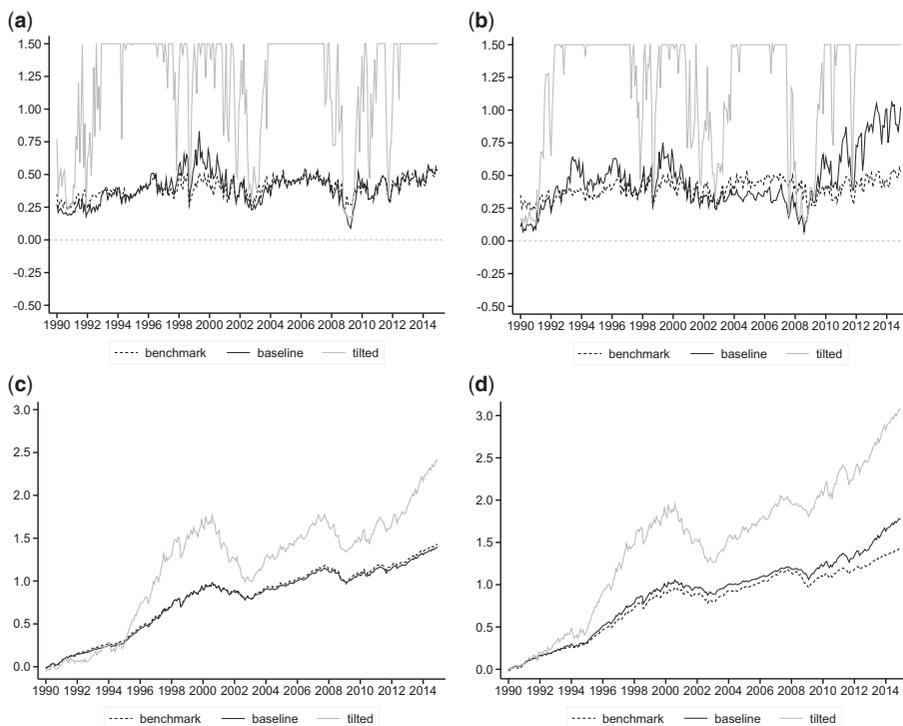
Figure 5. Asset allocation weights and cumulative wealth, SV model.

Notes: Panels (**a**) and (**b**) show the time series of equity weights in Equation (33) of the monthly port-folios for the earnings-price (EP) ratio and the short-interest index (SII) baseline and tilted densities, along with the equity weights from the historical-average with SV (HA–SV) benchmark. We compute the optimal allocation to stocks and T-bills based on the predictive density of excess returns. The in-vestor is assumed to have power utility with a coefficient of relative risk aversion $A = 3$ in Equation (30), while the equity weights are constrained to lie in the $[-0.5, 1.5]$ interval. Panels (**c**) and (**d**) show the corresponding log cumulative wealth. All forecasts are OOS using recursive estimates for 1990:01–2014:12.

In the case of EP (SII), the baseline weights are between 0.090 (0.070) and 0.830 (1.060), with an average of 0.397 (0.456). The tilted weights are between 0.170 (0.130) and 1.500 (1.500) with an average of 1.199 (1.189) for EP (SII). The fact that the tilted weights are generally larger than the baseline and benchmark ones means that the tilted densities tend to imply larger equity positions. The bottom panels of the same figure show the corresponding log cumulative wealth for the three portfolios, computed using Equation (34). By and large, the wealth generated by the tilted density forecasts lies above its baseline and benchmark counterparts, a pattern that is consistent throughout the whole OOS period, and in line with the CERs reported in Table 2.

4 Robustness Checks

In this section, we summarize the results of two robustness checks we performed to validate the empirical results presented in Sections 2 and 3 regarding the SV model. First, we explore the robustness of our results to only tilting the volatility of the baseline density forecasts,

Table 4. Robustness check I: altering the second moment only, SV model

Model	ALSD		CER gains, $A = 3$		CER gains, $A = 5$	
	(1) Baseline	(2) Tilted	(3) Baseline	(4) Tilted	(5) Baseline	(6) Tilted
DP	−0.008	**0.199*****	−0.514	**1.171**	−0.319	**0.516**
DY	−0.009	**0.198*****	−0.510	**1.038**	−0.302	**0.415**
EP	−0.003	**0.207*****	−0.155	**1.783**	−0.082	**1.987**
DE	−0.011	**0.189*****	−0.061	**2.484**	−0.033	**1.474**
RVOL	−0.010	**0.193*****	−0.098	**0.979**	−0.060	**0.476**
BM	0.000	**0.205*****	0.111	**1.789**	0.086	**1.550**
NTIS	−0.011	**0.187*****	−0.849	**1.393**	−0.494	**0.374**
TBL	0.003	**0.199*****	0.750	**2.530**	0.453	**1.857**
LTY	0.001	**0.207*****	0.426	**1.848**	0.284	**1.797**
LTR	−0.002	**0.210*****	−0.099	**1.575**	−0.037	**1.535**
TMS	−0.003	**0.190*****	0.216	**4.124**	0.161	**2.500**
DFY	−0.010	**0.197*****	−0.747	**0.153**	−0.432	−0.243
DFR	−0.005	**0.192*****	−0.104	**3.709**	−0.018	**2.979**
INFL	0.001	**0.209*****	0.129	**1.890**	0.091	**1.913**
SII	0.005	**0.216*****	1.405	**4.923**	0.849	**4.221**
KS	−0.029	**0.119*****	0.726	0.540	0.252	−0.065
EWC	−0.003	**0.211*****	0.106	**2.598**	0.066	**2.431**

Notes: The table shows results for the OOS statistical and economic predictability of the SV model tilting the baseline distributions by altering their second moment only for 1990:01–2014:12. Bold numbers indicate all instances where the tilted forecasts improve upon the corresponding baseline forecasts. *,**,***Statistical significance at 10, 5 and and 1% levels respectively, using the DM (1995) tests discussed in Section 3.1.

without imposing the non-negativity constraint on its mean. Next, we test the robustness of our main results to the use of an alternative benchmark model.[24]

4.1 Robustness I: Tilting Altering Volatility Only

Our main results for the SV model in Section 3 were obtained by altering both the first and second moments of the baseline predictive density $p(r_t|\mathcal{D}^{t-1})$. To isolate how much of the improvement in forecast performance we found stems from the forward-looking information in options prices alone, Table 4 presents results of our tilting procedure when we only alter the second moment of the baseline densities, without imposing the non-negativity constraint on their means. Note that we omit the results for the R^2_{OOS} statistics, as in this particular case, by altering only the volatility of the predictive densities, the point forecasts from the baseline and tilted densities are identical.

24 We should also point out that while throughout our analysis we have focused on a 1-month forecast horizon, our approach can be extended to forecast horizons of more than 1 period (1 month) in two alternative ways. The first is to use the 1-month VIX to tilt density forecasts for longer horizon returns, which hinges on the assumption that options with 1 month to expiration can be used to predict returns at longer horizons as in Bollerslev, Tauchen, and Zhou (2009)—see their Section 3. The second is to construct risk neutral measures of volatility using options with expiration matching the forecast horizon, keeping in mind options data availability.

Table 5. Robustness check II: SV model with GARCH(1,1) as benchmark

Model	OOS R^2		ALSD		CER gains, $A = 3$		CER gains, $A = 5$	
	(1) Baseline	(2) Tilted	(3) Baseline	(4) Tilted	(5) Baseline	(6) Tilted	(7) Baseline	(8) Tilted
DP	−0.065	**0.013**	−0.168	**0.040***	0.382	**2.239**	0.216	**1.245**
DY	−0.100	**0.043**	−0.169	**0.039***	0.385	**2.355**	0.233	**1.284**
EP	0.429	**0.616**	−0.163	**0.047***	0.741	**2.726**	0.453	**2.594**
DE	−0.487	**0.047**	−0.171	**0.029**	0.835	**3.514**	0.503	**2.118**
RVOL	0.372	**0.784***	−0.171	**0.041***	0.797	**2.213**	0.475	**1.305**
BM	0.395	0.372	−0.160	**0.045***	1.006	**2.623**	0.622	**2.031**
NTIS	−1.962	**−1.220****	−0.171	**0.035**	0.046	**2.890**	0.042	**1.473**
TBL	0.322	0.254	−0.157	**0.039**	1.646	**3.360**	0.988	**2.366**
LTY	0.759	0.753	−0.159	**0.046***	1.322	**2.730**	0.819	**2.301**
LTR	0.808	0.721	−0.162	**0.049****	0.797	**2.388**	0.499	**2.017**
TMS	0.380	0.349	−0.164	**0.030**	1.111	**4.995**	0.696	**3.047**
DFY	−0.280	**0.128***	−0.170	**0.042***	0.149	**2.047**	0.103	**1.156**
DFR	0.272	−0.451	−0.166	**0.032**	0.791	**4.580**	0.517	**3.469**
INFL	0.006	**0.052**	−0.159	**0.049***	1.025	**2.880**	0.626	**2.522**
SII	2.557	**2.583**	−0.155	**0.056****	2.301	**5.786**	1.384	**4.734**
KS	−12.494	**−3.702**	−0.189	−1.868	1.621	**3.040**	0.787	**1.983**
EWC	1.012	1.012	−0.163	**0.051****	1.002	**3.605**	0.602	**2.992**

Notes: The table shows results for the OOS statistical and economic predictability of the SV model using the HA–GARCH benchmark for 1990:01–2014:12. Bold numbers indicate all instances where the tilted forecasts improve upon the corresponding baseline forecasts. ***Statistical significance at 10, 5 and 1% levels respectively, using the DM (1995) tests discussed in Section 3.1.

Starting with Column (2) of Table 4, we see that tilting the baseline SV densities in this way leads to essentially the same ALSDs that we obtained when tilting jointly the first two moments of the baseline predictive densities. The only exception is the KS specification, for which the tilted ALSD is not positive. In particular, a comparison between the entries of Column (2) in the bottom panel of Table 1 and the entries of Column (2) in Table 4 reveals that the differences in the ALSDs between the two tables do not exceed 0.008 in absolute value once we exclude the KS specification.

Similarly, a comparison between Column (2) in Table 2 and Columns (4) and (6) of Table 4 confirms that the CER gains of the two tilting procedures are very similar. More specifically, it appears that altering both moments (as opposed to only tilting the second moment of the predictive densities) only marginally improves CER gains. The only real exception to this pattern is for the KS specification, where we see that the non-negativity constraint on the mean leads to additional CER gains of 1.604% and 1.513% in the case of $A = 3$ and $A = 5$, respectively.

4.2 Robustness II: GARCH(1,1) Benchmark

Table 5 shows the economic and statistical predictability results associated with the baseline and tilted SV densities when the benchmark model is the HA–GARCH. Starting with the R^2_{OOS} values in Column (1) of Table 5, we note that the baseline SV density forecasts outperform the benchmark GARCH in 11 out of the 17 models. However, the DM tests

indicate that the SV baseline point forecasts are not significantly better than the benchmark ones at conventional levels. Turning to Column (2) of the table, it appears that tilting the baseline SV densities by jointly altering their volatility and mean produces positive R^2_{OOS} in 14 cases. In 3 out of these 14 cases, the resulting improvements are significant at conventional levels. Moving to the ALSDs results in Columns (3) and (4) of Table 5, we find that with the exception of the KS specification, the tilted SV density forecasts tend to be more accurate than those from the benchmark GARCH specification.

Table 5 delivers two important messages in terms of economic predictability. First, the baseline SV density forecasts consistently outperform the benchmark GARCH model. Second, tilting further improves the economic predictability of the baseline density forecasts, relative to the benchmark GARCH specification. For example, in the case of $A = 3$, the baseline CERDs are between 0.046% for net equity expansion (NTIS) and 2.301% for SII, with an average of 0.939%. The tilted CERDs are between 2.047% for the DFY and 5.786% for SII, with an average of 3.175%. In the case of $A = 5$, the range of the baseline CERDs is 0.042% (NTIS) to 1.384% (SII) with an implied average of 0.563%. The range for the tilted CERDs is from 1.156% (DFY) to 4.734% (SII) implying an average of 2.273%.

Finally, we are now in a position to assess the relative performance of the SV and GARCH models by comparing the appropriate columns of Tables 3 and 5. Using R^2_{OOS} as our metric, the tilted SV density forecasts outperform their GARCH(1,1) analogs for 8 out of the 17 models considered. Using ALSD, as our metric, the SV density forecasts are superior for eight models, too. In terms of economic predictability, the tilted SV density forecasts outperform the tilted GARCH(1,1) ones in 13 out of 17 models with average CERDs of 3.175% ($A = 3$) and 2.273% ($A = 5$) compared with 2.369% ($A = 3$) and 1.741% ($A = 5$). Hence, although the SV and GARCH(1,1) models perform equally well in terms of statistical criteria, the SV model performs better in terms of economic criteria.

5 Conclusions

The paper introduces a novel approach to refine density forecasts for the equity premium using information from the derivatives markets in a time-series setting. We tilt predictive densities from SV and GARCH models toward the second moment of the distribution implied by options prices, while imposing a non-negativity constraint on the mean of the resulting density. Tilting augments the *backward-looking* information in the baseline models with *forward-looking* information from the options in a straightforward manner that is not computationally intensive.

Using monthly density forecasts for the S&P500 between 1990 and 2014, we show that tilting significantly improves both the statistical and economic predictability of stock returns. Although improvements in forecasting the equity premium using information from the derivative markets have been previously documented, they have been limited to point forecasts, incorporating option-implied moments among predictors in forecasting regressions. Extending our method to higher moments, such as skewness and kurtosis, which have been receiving increased attention in empirical asset pricing, as well as longer investment horizons is a research agenda worth pursuing, especially as more options data become available.

Supplementary Data

References

Altavilla, C., R. Giacomini, and R. Constantini. 2014. Bond Returns and Market Expectations. *Journal of Financial Econometrics* 12: 708–729.

Altigan, Y., T. G. Bali, and K. O. Demitras. 2015. Implied Volatility Spreads and Expected Market Returns. *Journal of Business & Economic Statistics* 33: 87–101.

Amisano, G., and R. Giacomini. 2007. Comparing Density Forecasts via Weighted Likelihood Ratio Tests. *Journal of Business & Economic Statistics* 25: 177–190.

Andersen, T. G., T. Bollerslev, P. F. Christoffersen, and F. X. Diebold. 2006. "Volatility and Correlation Forecasting." In G. Elliott, C. W. J. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*, vol. 1. pp. 777–878. Amsterdam, The Netherlands: North Holland.

Back, K. 2010. *Asset Pricing and Portfolio Choice Theory*. New York: Oxford University Press.

Barberis, N. 2000. Investing for the Long Run When Returns Are Predictable. *Journal of Finance* LV: 225–264.

Bauwens, L., and M. Lubrano. 1998. Bayesian Inference on GARCH Models Using the Gibbs Sampler. *Econometrics Journal* 1: C23–C46.

Bloom, N. 2009. The Impact of Uncertainty Shocks. *Econometrica* 77: 623–685.

Bollerslev, T., G. Tauchen, and H. Zhou. 2009. Expected Stock Returns and Variance Risk Premia. *Review of Financial Studies* 22: 4463–4492.

Campbell, J. Y., and S. Thompson. 2008. Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *Review of Financial Studies* 21: 1509–1531.

Campbell, J. Y., and L. M. Viceira. 2002. *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*. New York: Oxford University Press.

Cenesizoglu, T., and A. Timmermann. 2012. Do Return Prediction Models Add Economic Value? *Journal of Banking and Finance* 36: 2974–2987.

Chan, J. C. C., and A. L. Grant. 2016. Modeling Energy Price Dynamics: GARCH Versus Stochastic Volatility. *Energy Economics* 54: 182–189.

Clark, T. E. 2011. Real-Time Density Forecasts from Bayesian Vector Autoregressions with Stochastic Volatility. *Journal of Business & Economic Statistics* 29: 327–341.

Clark, T. E., and M. McCracken. 2011. "Nested Forecast Model Comparisons: A New Approach to Testing Equal Accuracy." Federal Reserve Bank of St. Louis Working Paper.

Clark, T. E., and F. Ravazzolo. 2015. Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility. *Journal of Applied Econometrics,* 30: 551–575.

Cogley, T., S. Morozov, and T. J. Sargent. 2005. Bayesian Fan Charts for U.K. Inflation: Forecasting and Sources of Uncertainty in an Evolving Monetary System. *Journal of Economic Dynamics & Control* 29: 1893–1925.

Diebold, F. X., and R. S. Mariano. 1995. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics* 13: 253–263.

Engle, R. F., and G. Gallo. 2006. A Multiple Indicators Model for Volatility Using Intra-daily Data. *Journal of Econometrics* 131: 3–27.

Faust, J., and J. H. Wright. 2013. "Forecasting Inflation." In G. Elliott and A. Timmermann (eds.), *Handbook of Economic Forecasting*, vol. 2, pp. 2–56. Amsterdam, The Netherlands: North Holland.

Giacomini, R., and G. Ragusa. 2014. Theory-Coherent Forecasting. *Journal of Econometrics* 182: 145–155.

Giacomini, R., and B. Rossi. 2010. Forecast Comparisons in Unstable Environments. *Journal of Applied Econometrics* 25: 595–620.

Goyal, A., and I. Welch. 2008. A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies* 21: 1455–1508.

Hansen, P. R., Z. Huang, and H. H. Shek. 2012. Realized GARCH: A Joint Model for Returns and Realized Measures of Volatility. *Journal of Applied Econometrics* 27: 877–906.

Harvey, D., S. Leybourne, and P. Newbold. 1997. Testing the Equality of Prediction Mean Squared Errors. *International Journal of Forecasting* 13: 281–291.

Heston, S., and S. Nandi. 2000. A Closed-Form GARCH Option Valuation Model. *Review of Financial Studies* 13: 585–625.

Johannes, M. S., A. Korteweg, and N. Polson. 2014. Sequential Learning, Predictive Regressions, and Optimal Portfolio Returns. *Journal of Finance* 69: 611–644.

Jorion, P. 1995. Predicting Volatility in the Foreign Exchange Market. *Journal of Finance* 50: 507–528.

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

Koop, G. 2003. *Bayesian Econometrics*. West Sussex, England: John Wiley & Sons, Ltd.

Krüger, F., T. Clark, and F. Ravazzolo. 2015. Using Entropic Tilting to Combine BVAR Forecasts with External New Casts. *Journal of Business & Economic Statistics* 35: 470–485.

Pettenuzzo, D., A. Timmermann, and R. Valkanov. 2014. Forecasting Stock Returns under Economic Constraints. *Journal of Financial Economics* 114: 517–553.

Poon, S.-H., and C. W. J. Granger. 2003. Forecasting Volatility in Financial Markets: A Review. *Journal of Economic Literature* 41: 478–539.

Primiceri, G. E. 2005. Time Varying Structural Vector Autoregressions and Monetary Policy. *The Review of Economic Studies* 72: 821–852.

Rapach, D. E., M. C. Ringgenberg, and G. Zhou. 2016. Short Interest and Aggregate Stock Returns. *Journal of Financial Economics*, 121: 46–65.

Rapach, D. E., and G. Zhou. 2013. "Forecasting Stock Returns." In G. Elliot and A. Timmermann (eds.), *Handbook of Economic Forecasting*, vol. 2, pp. 329–383. Amsterdam, The Netherlands: North Holland.

Robertson, J. C., E. W. Tallman, and C. H. Whiteman. 2002. "Forecasting Using Relative Entropy." Federal Reserve Bank of Atlanta WP 2002-22.

Robertson, J. C., E. W. Tallman, and C. H. Whiteman. 2005. Forecasting Using Relative Entropy. *Journal of Money, Credit and Banking* 37: 383–401.

Shephard, N., and K. Sheppard. 2010. Realising the future: Forecasting with High-Frequency-Based Volatility (Heavy) Models. *Journal of Applied Econometrics* 25: 197–231.

Szakmary, A., E. Ors, J. K. Kim, and W. N. Davidson. 2003. The Predictive Power of Implied Volatility: Evidence from 35 Futures Markets. *Journal of Banking and Finance* 27: 2151–2175.

Young Chang, B., P. Christoffersen, and K. Jacobs. 2013. Market Skewness Risk and the Cross Section of Stock Returns. *Journal of Financial Economics* 107: 46–68.

Zivot, E. 2009. *Practical Issues in the Analysis of Univariate GARCH Models*, pp. 113–155. Berlin, Germany: Springer.