

Option-Implied Equity Premium Predictions via Entropic Tilting

Konstantinos Metaxoglou Davide Pettenuzzo Aaron Smith *

January 4, 2016

Abstract

We propose a new method to improve density forecasts of the equity premium using information from options markets. We tilt the predictive densities from standard econometric models suggested in the stock return predictability literature towards the second moment of the risk-neutral distribution implied by options prices. In so doing, we use a simple regression-based approach to remove the variance risk premium. By combining the *backward-looking* information contained in the econometric models with the *forward-looking* information from the options prices, tilting yields sharper predictive densities. Using density forecasts of the U.S. equity premium in [Rapach and Zhou \(2012\)](#), we find that tilting leads to more accurate predictions, both in terms of statistical and economic performance.

JEL classification: C11, C22, G11, G12.

Keywords: entropic tilting, density forecasts, variance risk premium, equity premium, options.

*Metaxoglou: Department of Economics, Carleton University, konstantinos.metaxoglou@carleton.ca.
Smith: Department of Agricultural and Resource Economics, University of California, Davis, adsmith@ucdavis.edu.
Pettenuzzo: Brandeis University, dpettenu@brandeis.edu.

1 Introduction

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. Derivative prices convey information about the conditional density of future outcomes and, hence, they are inherently *forward looking*. They contain information about market expectations and should be useful for improving return forecasts.

In this paper, we provide a simple procedure to blend backward- and forward-looking information to sharpen predictive densities of the equity premium obtained from existing econometric models. Our approach entails taking a given predictive density and tilting it towards moments implied by options prices. Specifically, we extract the variance of the risk-neutral distribution from options prices and subtract from it a regression-based estimate of the variance risk premium to obtain a forward-looking variance estimate. We then twist the baseline physical predictive distribution towards this forward-looking variance using entropic tilting. The procedure is simple; using a few lines of code, we modify the predictive density such that its variance resembles the risk-neutral variance with the variance risk premium removed.

To illustrate our method, we use the equity premium on the S&P 500. We take the models in [Rapach and Zhou \(2012\)](#) and tilt their implied predictive densities towards the option-implied variance. We find that entropic tilting significantly improves the out-of-sample predictability of stock returns, both in terms of statistical and economic measures of forecasting accuracy. Improvements in forecasting performance using option-implied information have been well documented ([Christoffersen et al. \(2013\)](#)). Christoffersen et al. find that combining historical information with information extracted from derivative prices can improve forecasts significantly. Numerous papers also show that option-implied volatility provides a good forecast

of future realized volatility (e.g., [Szakmary et al. \(2003\)](#)) and predicts the equity premium (e.g., [Bollerslev et al. \(2009\)](#)).

We make two contributions to the literature on improving forecasts using option-implied information. 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 focuses on point forecasts and incorporates option-implied moments among the predictors in a forecasting regression. (e.g., [Altigan et al. \(2015\)](#)). Furthermore, our method can be extended to higher moments, such as skewness and kurtosis, which have received increased attention recently in empirical asset pricing overall ([Young Chang et al. \(2013\)](#)).

The remainder of the paper is organized as follows. In [Section 2](#), we first provide an overview of entropic tilting and a method to construct predictive densities for the equity premium. Our approach for removing the variance risk premium from the variance of the risk-neutral distribution implied by option prices follows. [Section 3](#) presents our main results, and [Section 4](#) focuses on out-of-sample statistical and economic performance. Finally, we provide some concluding remarks. All tables and figures are provided after the main body of the text.

2 Entropic Tilting for Equity Premium Forecasting

Entropic tilting is a highly flexible non-parametric method to change the shape of a baseline predictive distribution to incorporate additional information about an underlying random variable of interest. Such additional information may come in the form of moments, and this is the approach we follow here. In particular, we use the density forecasts implied by the model combination approach of [Rapach and Zhou \(2012\)](#) as baseline distributions. We use entropic tilting to incorporate moment restrictions derived from options prices in

these baseline distributions. In what follows, we outline the general tilting method, how we incorporate the moment-based information from the options markets into the baseline density forecasts, and the econometric model we rely on to produce the baseline density forecasts for the equity premium.

2.1 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 . 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 $g(r_{t+1})$ defined such that

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

For example, $g(r_{t+1})$ may represent quantities such as the mean, $g(r_{t+1}) = r_{t+1}$, the variance, $g(r_{t+1}) = (r_{t+1} - E[r_{t+1}])^2$, or higher moments of the predictive distribution; see e.g., [Robertson et al. \(2005\)](#) 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 \(2013\)](#), or could be coming from survey forecasts and model-based nowcasts as in [Altavilla et al. \(2014\)](#) and [Krüger et al. \(2015\)](#).

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

$$\int g(r_{t+1})p(r_{t+1}|\mathcal{D}^t) dr_{t+1} \neq \bar{g}_t. \quad (2)$$

Thus, by transforming $p(r_{t+1}|\mathcal{D}^t)$ so that (1) holds, we sharpen 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 but that satisfies the moment restriction implied by (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 density¹

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

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

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

where the last constraint may be viewed as the Monte-Carlo approximation to the moment restriction in (1) using the language in Cogley et al. (2005).² The implied first-order conditions are given by

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

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

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

As a result, the baseline weights are *tilted* in an exponential fashion via (6) to generate the

¹Other measures of divergence could be chosen. As Giacomini and Ragusa (2013) note, the two main reasons for working with the Kullback-Leibler measure are that: (i) it provides a convenient analytical expression for the tilted weights; (ii) 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 (see e.g., Amisano and Giacomini (2007)).

²See Robertson et al. (2005) and the references there in. 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.

new weights. The tilting parameter γ_t^* can be found by solving the minimization problem

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

In our case, we use the variance of the risk-neutral distribution for the equity premium, as implied by the option markets, to distort the baseline predictive distribution $p(r_{t+1}|\mathcal{D}^t)$ so that its dispersion, as captured by $Var(r_{t+1}|\mathcal{D}^t)$, resembles that of the option implied risk-neutral distribution. It is the forward-looking aspect of the options market that serves as the source of new information and is also the novelty in our approach. In the spirit of Robertson et al., we incorporate restrictions that could in principle be built directly into the forecasting model in a manner that is less demanding computationally.

2.2 Equity Premium Predictions

We now describe the econometric model we rely on to construct the baseline predictive density $p(r_{t+1}|\mathcal{D}^t)$. We follow [Rapach et al. \(2010\)](#) and [Rapach and Zhou \(2012\)](#), who find that model combinations based on popular predictors in the literature lead to significant improvements in the out-of-sample predictability of stock returns. In particular, Rapach and Zhou start with a set of linear univariate regression models, each including a constant and one of fourteen predictors described in [Goyal and Welch \(2008\)](#)

$$r_{t+1} = \mu_i + \beta_i x_{i,t} + \varepsilon_{i,t+1} \quad i = 1, \dots, K, \quad (8)$$

where $K = 14$, $x_{i,t}$ is the i -th predictor at time t , and $\varepsilon_{i,t+1} \sim \mathcal{N}(0, \sigma_i^2)$. The model in (8) entails estimating two mean parameters, μ_i and β_i , via an OLS regression. For each predictor $i \in \{1, \dots, K\}$ and time period t , we obtain the point forecast at time $t + 1$ conditional on

information available up to time t as follows³

$$\hat{r}_{i,t+1|t} = \hat{\mu}_{i,t} + \hat{\beta}_{i,t}x_{i,t}, \quad (9)$$

where $\hat{\mu}_{i,t}$ and $\hat{\beta}_{i,t}$ are the OLS estimators of (8) given by

$$\begin{pmatrix} \hat{\mu}_{i,t} \\ \hat{\beta}_{i,t} \end{pmatrix}' = \left(\sum_{\tau=2}^t z_{i,\tau-1} z_{i,\tau-1}' \right)^{-1} \left(\sum_{\tau=2}^t z_{i,\tau-1} r_{\tau} \right), \quad (10)$$

and $z_{i,\tau-1} \equiv (1, x_{i,\tau-1})'$. Next, Rapach and Zhou show that combining the individual forecasts in (9) delivers substantial improvements in out-of-sample forecasting accuracy. In one case, they combine point forecasts with weights that depend on recent forecasting performance on the basis of a discounted mean squared forecast error (DMSFE). More specifically, their DMSFE return forecast is given by

$$\hat{r}_{t+1|t}^{DMSFE} = \sum_{i=1}^K w_{i,t}^* \hat{r}_{i,t+1|t}, \quad (11)$$

where

$$w_{i,t}^* = \frac{\phi_{i,t}^{-1}}{\sum_{i=1}^K \phi_{i,t}^{-1}}, \quad \phi_{i,t} = \sum_{s=m}^{t-1} \theta^{t-1-s} (r_{s+1} - \hat{r}_{i,s+1|s})^2, \quad \theta = 0.75$$

with $m+1$ being the beginning of the out-of-sample period and θ being a discount factor.

We proceed in a similar fashion, but given that our goal is to produce a full predictive density for the equity premium, we use the DMSFE weights to construct a combined density forecast as follows

$$p(r_{t+1} | \mathcal{D}^t) = \sum_{i=1}^K w_{i,t}^* \times p(r_{t+1} | \mathcal{M}_i, \mathcal{D}^t), \quad (12)$$

³We use an expanding window approach, where we recursively extend the estimation window one month at a time to obtain new predictive coefficients and to construct an out-of-sample density forecast for the following month. We proceed in this way until the full sample is exhausted. The procedure mimics the real-time information set for an investor.

where $r_{t+1}|\mathcal{M}_i, \mathcal{D}^t \sim \mathcal{N}(\hat{r}_{i,t+1|t}, \hat{\sigma}_i^2)$, \mathcal{M}_i identifies the univariate model based on predictor i , and $\hat{\sigma}_i^2$ denotes the OLS estimates of the corresponding variance.

2.3 Removing the Variance Risk Premium

We capitalize on the literature that has demonstrated the predictive power of implied volatility regarding 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.

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 et al. \(2009\)](#) provides 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 $p(r_{t+1}|\mathcal{D}^t)$ towards it.

Let $\hat{u}_{\mathbb{P},t+1}$ denote the forecast error from the baseline physical predictive distribution at time $t + 1$ obtained following the approach in the previous section

$$\hat{u}_{\mathbb{P},t+1} = r_{t+1} - E(r_{t+1}|\mathcal{D}^t), \quad (13)$$

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

VRP_{t+1}

$$\sigma_{\mathbb{P},t+1}^2 = \sigma_{\mathbb{Q},t+1}^2 - VRP_{t+1}. \quad (14)$$

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

$$\log(\sigma_{\mathbb{P},t+1}^2) = \alpha + \beta \log(\sigma_{\mathbb{Q},t+1}^2). \quad (15)$$

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

$$\hat{\sigma}_{\mathbb{P},t+1}^2 = \exp\left(\hat{\alpha} + \hat{\beta} \log(\sigma_{\mathbb{Q},t+1}^2)\right), \quad (16)$$

which implies that the variance risk premium is

$$\begin{aligned} VRP_{t+1} &= \sigma_{\mathbb{Q},t+1}^2 - \exp\left(\alpha + \beta \log(\sigma_{\mathbb{Q},t+1}^2)\right) \\ \widehat{VRP}_{t+1} &= \sigma_{\mathbb{Q},t+1}^2 - \exp\left(\hat{\alpha} + \hat{\beta} \log(\sigma_{\mathbb{Q},t+1}^2)\right). \end{aligned} \quad (17)$$

To implement our approach, for our measure of risk-neutral variance of the S&P 500 returns, $\sigma_{\mathbb{Q},t+1}^2$, we follow [Bollerslev et al. \(2009\)](#) and use the end-of-month values of the squared Chicago Board Options Exchange (CBOE) Volatility Index (VIX). In 1993, the CBOE introduced VIX, originally designed to measure the market's expectation of 30-day volatility implied by ATM S&P 100 Index (OEX) option prices. In 2003, CBOE together with Goldman Sachs updated the methodology and formula for VIX. The new VIX is based on the S&P 500 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.

Figure 1 provides a time series plot of the end-of-month values of the VIX squared between January 1995 and December 2010—January 1995 is the first month in our sample for which we tilt the baseline predictive densities. Setting aside the very prominent spike in October 2008 around the peak of the most recent financial crisis, our measure of risk-neutral variance is higher during 1997–2003, a period of well-documented turmoil in the 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).

3 Empirical Results

We obtain the data necessary to generate the predictive densities in (12) from Goyal and Welch (2008), who subsequently extended their original data to December 2010.⁴ End-of-month stock returns are computed from the S&P500 index and include dividends. A short T-bill rate is subtracted from stock returns in order to obtain the excess returns shown in Figure 2. The red and green vertical lines in the figure indicate the first month for which the VIX is available, January 1990, and the first month for which we tilt the baseline predictive density, January 1995. When annualized, the equity premium has an average of 5.6% and a standard deviation of 19.3% over the period January 1926 to December 2010. During the more recent period, January 1995 to December 2010, the mean and standard deviation of the equity premium are 5.1% and 16.1%, respectively. In this later part of the sample, we can easily identify the highly-volatile periods of 1997–2003 and the recent financial crisis through Figure 2.

In the empirical exercises that follow, we begin by computing the baseline predictive densities for the equity premium using (12). For the univariate models' densities, $p(r_{t+1}|\mathcal{M}_i, \mathcal{D}^t)$, $i = 1, \dots, K$, we start in January 1980 and proceed in a recursive fashion using an expanding-

⁴The data are readily available from Amit Goyal's website at <http://www.hec.unil.ch/agoyal/>.

window approach until the last observation in the sample. Hence, we produce K time series of one-step-ahead density forecasts between January 1980 and December 2010. Next, we combine the univariate models' densities into $p(r_{t+1}|\mathcal{D}^t)$, holding out the period January 1980 to December 1989 to train the combination weights. This leaves us with an initial out-of-sample period from January 1990 to December 2010. Subsequently, we tilt the DMSFE predictive densities towards the option implied risk-neutral variance using (16) to remove the variance risk premium. In particular, we obtain our estimates for the variance risk premium by regressing the log squared forecast error implied by the DMSFE predictive density on the the log squared VIX using (15) and an expanding-window approach. Setting aside the period January 1990 to December 1994 to estimate the first variance risk premium, our final out-of-sample period is January 1995 to December 2010.

3.1 Variance Risk Premium

The top panel of Figure 3 shows three different variance measures that are relevant in our exercise. The first is the variance associated with the baseline DMSFE predictive densities (baseline) in (12). The second is the end-of-month squared VIX, which is our measure of the risk-neutral variance (risk-neutral). The third is the variance associated with the tilted DMSFE predictive densities (tilted). A striking feature of the baseline variance is that it exhibits very little variation over time, even during periods of well-documented turmoil. In addition, for most of the sample, the tilted variance is smaller than the baseline variance. The most notable exception is 2009, when the variance of the titled densities is larger than the variance of the baseline densities in 7 of the 12 months.

The bottom panel of Figure 3 provides the slope estimate from the variance risk premium regression in (16), along with 95% confidence intervals. The regressions use an expanding window with initial size of 60 monthly observations. The slope estimate fails to be statistically different from 1 during the entire 16-year period, which implies that the variance risk

premium equals approximately the average difference between the squared forecast error and the squared VIX.

[Figure 4](#) plots the evolution of the estimated variance risk premium in the out-of-sample period. Our estimated variance risk premium is high during the periods of economic uncertainty, such as the recent financial crisis (Fall 2008), around the WorldCom and Enron scandals (Summer/Fall 2002) and Gulf War II (Spring 2003), as well as close the Russian financial crisis (Fall 1998). In addition, the same figure shows that our variance risk premium is highly comparable to the variance risk premium from [Bollerslev et al. \(2009\)](#) with the two series tracking each other closely. The only notable deviation between the two variance risk premium series is in October 2008. Bollerslev et al. compute their volatility risk premium by taking the difference between implied and realized volatility, whereas we obtain ours from a regression of the log squared forecast error on log squared VIX, so the similarity in these series provides some evidence that our log linear approximation is reasonable. Of course, the true test of our procedure is its forecasting performance.

3.2 Entropic Tilting

[Figure 5](#) shows the posterior probability intervals for the tilted and the baseline densities over the out-of-sample period. Consistent with our earlier findings, the shape of the tilted predictive densities exhibits much more variation over time compared to its baseline counterpart. For example, if we focus on the far left tail of the distributions, the 1% quantile for the baseline density forecasts is between -0.132 and -0.124, while that for the tilted density forecasts is between -0.262 and -0.033. In the case of the far right tail of the distributions, the 99% quantile for the baseline density forecasts is 0.131–0.138, while that for the tilted density forecasts is 0.043–0.265. Similar conclusions are drawn by looking at the shoulders of the two distributions.

As we discussed earlier in paper, the empirical KLIC defined in [\(3\)](#) gauges how much the

baseline density is altered by the tilting procedure. That is, small values of the empirical KLIC signify agreement between the baseline predictive model and outside information, while large values signify disagreement.⁵ 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. The average KLIC for the entire out-of-sample period is 0.21, which is within the range of KLIC values reported in Cogley et al. (2005) and Robertson et al. (2002), 0.12–0.68 and 0.06–0.66, respectively.⁶ 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 et al.—these values can serve as benchmark for aggressive twisting given that the consumption-CAPM is known to fit the data poorly. In our case, although the KLIC achieves some of its largest value in 1995–1996, 2004–2007, and also in 2009, its annual averages never exceed 0.6 in any of these years. In summary, it appears that for the largest part of out sample, the twisting of the baseline distributions is not excessively aggressive.

To conclude this section, Figure 6 gives a sense of the magnitude of the tilting effect on the baseline predictive distributions at four distinct dates. In each of these panels, we provide kernel plots for both the baseline and tilted predictive densities, along with a vertical line indicating the realized value of the equity premium on that date. In the top two panels—referring to December 1997 and August 1999, respectively—tilting has a very moderate effect on the baseline distributions, which are already properly centered with respect to the observed value of the equity premium. In the third panel (January 2006), tilting sharpens

⁵Note that when $KLIC(\pi_t^*; \pi)$ is zero, it means that the baseline and tilted densities coincide.

⁶The ranges reported here are based on Table 2 in Cogley et al., and on the KLIC statistics reported in Tables 1b, 2b, and 3b in Robertson et al..

the baseline density forecasts by reducing its variance.⁷ Finally, in the fourth panel (May 2009), tilting turns out to be detrimental to the baseline predictive density, as it shifts mass to the tails of the distribution and increasing the variance of the tilted density. Recall that 2009 is the year in which the variance of the tilted distribution is systematically higher than the variance of the baseline distribution.

4 Out-of-Sample Performance

In this section, we look into the question of whether the approach introduced in [Section 2](#) leads to more accurate equity premium forecasts, both in terms of statistical and economic criteria. The predictive accuracy is measured relative to the historical average (HA) as in [Goyal and Welch \(2008\)](#) and [Rapach and Zhou \(2012\)](#). However, we focus on density forecasts as opposed to point forecasts.

4.1 Specification Testing

The probability integral transformation (PIT) provides an alternative global assessment of the predictive density and can be complementary to the metrics discussed above. If the predictive density equals the true density, then the realized CDF of the predictive density will have a uniform distribution on the zero-to-one interval, i.e, for an optimal forecast CDF G_t , the PIT random variable ξ_t is *iid* $U(0, 1)$. A series of diagnostic checks can be applied on the PIT values or the values given by their inverse normal transformations $\tilde{\xi}_t = \Phi^{-1}(\xi_t)$, where $\Phi(\cdot)$ is the standard normal CDF. The basic premise here is that if $\{\xi_t\}_{t=m+1}^T \sim \text{iid } U(0, 1)$, then $\{\tilde{\xi}_t\}_{t=m+1}^T \sim \text{iid } N(0, 1)$ for a correctly specified predictive density.

The PITs can be tested for uniformity using the Kolmogorov-Smirnov (KS) tests or its

⁷For the vast majority of the forecast dates in our sample, the reshaping of the baseline distribution is very similar to what we see for January 2006.

Anderson-Darling (AD) modification—see Rossi and Sekhposyan (2014) and the references therein for a more detailed discussion of these tests. The tests measure the difference between the PIT and the CDF of a uniform distribution. The Doornik-Hansen (DH) test for normality of the $\tilde{\xi}$ series is based on the sum of the squares of transformed skewness and kurtosis measures (Doornik and Hansen (2008)).⁸

We also test for independence in the first and second central moments of the PITs using the Ljung-Box (LB)—Ljung and Box (1978)—test of serial correlation and a maximum lag length of 12 given that we use monthly data. Correct specification of density forecasts requires independence of the PITs. Under the maintained hypothesis of normality, using the Berkowitz (2001) likelihood ratio (LR) test, the joint null of goodness-of-fit (correct mean and variance) is tested against an AR(1) alternative with mean and variance possibly different from zero and one, respectively. An LR test with three degrees of freedom is based on $\delta = 0$, $\rho = 0$, and $s_\varepsilon = 1$, using the equation below

$$\tilde{\xi}_t - \delta = \rho(\tilde{\xi}_{t-1} - \delta) + \varepsilon_t, \quad \varepsilon_t \sim N(0, s_\varepsilon^2). \quad (18)$$

Testing the significance of ρ in (18) gives a test of the independence component alone, while testing for $\delta = 0$ and $s_\varepsilon = 1$ pertains to normality alone. On one hand, there is evidence that the tests based on the inverse normal of the PIT à la Berkowitz are more powerful than those based directly to the PIT, at least, in finite samples. On the other hand, detecting violations of the normality assumption only through the first two moments is a limitation.

In the first part of Table 1, we report the p-values of two PIT-based tests of uniformity; namely, the Kolmogorov-Smirnov (KS) and the Anderson-Darling (AD). Both the KS and the AD tests—under the maintained assumption of independence—favor the tilted over the baseline densities. In particular, the AD test, which has a power advantage over the KS

⁸Rossi and Sekhposyan highlight the power advantage of the AD over the KS test. They also provide MATLAB codes for implementing the tests described here at <http://www.tateviksekhposyan.org/>.

test, strongly rejects the null of correct specification for the baseline and HA densities.

In the absence of mis-specification, the PITs should be independent. We test for independence using the Ljung-Box test of no autocorrelation in the first (LB-I) and second (LB-II) central moments of the PITs. The p-values for the independence tests are reported in the second part of [Table 1](#). Although we fail to reject the null of no autocorrelation in both moments for the PITs associated with the tilted densities (p-values in excess of 0.44), we fail to reject the null only for the first central moment of the baseline and HA PITs, where we see p-values of 0.657 and 0.645, respectively. In addition, we strongly reject the null of no autocorrelation in the second moments of the baseline PITs.

In the last part of the table, we employ tests based on the inverse normal transformation of the PITs. Overall, there is strong evidence for mis-specification for the baseline and HA densities on the basis of the normality (Berkowitz I), as well as the normality-and-independence (Berkowitz III) tests. We strongly fail to reject the null for all three Berkowitz tests in the case of the tilted densities with p-values greater than or equal to 0.30. The Doornik-Hansen normality test, which relies on transformations of the third and fourth moments, points to mis-specification for all three densities.

4.2 Statistical Forecasting Performance

Predictive densities are collections of probabilities assigned to potential future outcomes. Hence, we use tools from the density forecasting evaluation literature to evaluate their relative accuracy. For the purpose of the discussion below, let T denote the total sample size and let $T - m$ be the number of out-of-sample observations. Following [Amisano and Giacomini \(2007\)](#), we consider the average log score differential

$$ALSD_{g,h} = \frac{1}{T - m} \sum_{t=m+1}^T LSD_{g,h}(t) \quad (19)$$

where $LSD_{g,h}(t) \equiv LS_{g,t} - LS_{h,t}$, and $LS_{g,t}$ ($LS_{h,t}$) denotes the time- t log predictive score for density g (h). For example, in our application, g and h may correspond to the baseline DMSFE and the historical-average (HA) predictive densities. Alternatively, g and h may correspond to the tilted DMSFE and the HA predictive densities. 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 $ALSD_{g,h}$ value indicates that, on average, g is more accurate than h in predicting the outcome of interest.

We also calculate the cumulative log-score difference (CLSD) between densities g and h , to track the relative ranking of the two predictive densities through the entire out-of-sample period

$$CLSD_{g,h}(t) = \sum_{\tau=m+1}^t LSD_{g,h}(\tau), \quad t = m+1, \dots, T. \quad (20)$$

If g is more accurate than h in every period out-of-sample, then the $CLSD_{g,h}$ line is monotonically increasing. Conversely, periods during which density g is less accurate than density h generate dips in the $CLSD_{g,h}$ line.

The top section of [Table 2](#) reports ALSDs for the tilted-HA and baseline-HA pairs based on [\(19\)](#). We do so for three distinct sub-samples, 1995–2000, 2001–2005, and 2006–2010, which are long enough to give us a fairly reliable estimate for the average LSDs. The superior performance of the tilted density forecasts, especially after 2000, is clear. The average LSD for the tilted-HA pair is around 0.2, while that for the baseline-HA pair never exceeds 0.003.⁹ We also employ a [Diebold and Mariano \(1995\)](#) one-sided test to assess whether the differences in forecast accuracy of the HA, baseline, and tilted densities over the entire out-of-sample period are statistically significant. The last row of [Table 2](#) reports Diebold-Mariano p-values of the t-statistic from regressing the LSDs on a constant over the whole out-of-sample

⁹Note also that there is a positive relationship between the LSD for the titled-HA pair and KLIC. The slope coefficient in an OLS regression of the LSD on the KLIC is positive and highly significant for both the entire sample and the three sub-samples.

period. Following [Clark and Ravazzolo \(2015\)](#), we compute the t-statistic using Newey-West standard errors with one lag and a pre-whitening of the LSDs.¹⁰ Based on the p-values of the Diebold-Mariano statistic, we strongly reject the null of equal predictive accuracy between the tilted and HA densities; the p-value for the statistic is less than 0.001. However, we fail to do so for the baseline and HA densities—the p-value of the statistic is 0.436.

[Figure 7](#) shows time-series plots of the cumulative log score difference (CLSDs), where, once again, we pit the baseline and tilted densities against the HA (panel (a)). The baseline DMSFE densities fail to beat the HA baseline between February 1995 and December 2008 with very few exceptions. The CLSD for the baseline-HA pair systematically exceeds zero only in 2009 and 2010, the last two years of our sample (panel (b)). On the other hand, the tilted predictive densities win the horse race against the HA densities by a large margin. In panel (b), the CLSD for the tilted densities exhibits an upward trend for January 1995 to June 1996 (period I), loses some ground between September 1998 and April 2003 (period II), and then continues with an upward trend up to May 2008 (period III), before it flattens out in the last part of the sample starting in early 2009 (period IV). For both periods I and III, the variance of the baseline density systematically exceeds its counterpart of the tilted density and the distance and the KLIC attains some of its largest values.

To sum up, the tilted DMSFE densities outperforms the baseline DMSFE densities for the equity premium on the basis of numerous statistics and diagnostics used in the density forecast evaluation literature.

4.3 Economic Performance

Up to this point, we compared the statistical performance of the baseline, tilted, and HA predictive densities. In this section, we compare their economic performance using the setup

¹⁰Citing Monte Carlo evidence in [Clark and McCracken \(2011\)](#), with nested models, Clark and Ravazzolo argue that the Diebold-Mariano 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.

of an investor utilizing these densities to make optimal portfolio decisions, taking parameter uncertainty into consideration as in [Kandel and Stambaugh \(1996\)](#) and [Barberis \(2000\)](#)), among others.

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.¹¹

Our interest lies on 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 into risky assets, and r_t being time t equity premium. The economic agent solves the optimal asset allocation problem

$$\omega_{t-1}^* = \arg \max_{\omega_{t-1}} E [U(\omega_{t-1}, r_t) | \mathcal{D}^{t-1}], \quad (21)$$

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

$$U(\omega_{t-1}, r_t) = \frac{[(1 - \omega_{t-1}) \exp(r_{f,t-1}) + \omega_{t-1} \exp(r_{f,t-1} + r_t)]^{1-A}}{1 - A}, \quad (22)$$

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 portfolio optimization problem using information available only at time $t - 1$. Taking expectations with respect to the predictive density of r_t , we can rewrite (21) as follows

$$\omega_{t-1}^* = \arg \max_{\omega_{t-1}} \int U(\omega_{t-1}, r_t) p(r_t | \mathcal{D}^{t-1}) dr_t. \quad (23)$$

The integral in (23) can be approximated using draws from the competing predictive den-

¹¹Sufficient conditions include quadratic utility or elliptical return distributions ([Back \(2010\)](#)).

sities. Specifically, using the HA predictive density h , we can approximate the solution to (23) using a large number of draws J , $\{r_{h,t}^j\}_{j=1}^J$, using the following expression

$$\hat{\omega}_{h,t-1} = \arg \max_{\omega_{t-1}} \frac{1}{J} \sum_{j=1}^J \left\{ \frac{[(1 - \omega_{t-1}) \exp(r_{f,t-1}) + \omega_{t-1} \exp(r_{f,t-1} + r_{h,t}^j)]^{1-A}}{1-A} \right\}. \quad (24)$$

Similarly, using $g \in \{base, tilt\}$ to denote the baseline and tilted DMSFE predictive densities, we can approximate (23) via

$$\hat{\omega}_{g,t-1} = \arg \max_{\omega_{t-1}} \frac{1}{J} \sum_{j=1}^J \left\{ \frac{[(1 - \omega_{t-1}) \exp(r_{f,t-1}) + \omega_{t-1} \exp(r_{f,t-1} + r_{g,t}^j)]^{1-A}}{1-A} \right\}. \quad (25)$$

The sequence of portfolio weights $\{\hat{\omega}_{h,t-1}\}_{t=m+1}^T$ and $\{\hat{\omega}_{g,t-1}\}_{t=m+1}^T$ are used to compute the realized utilities under the HA, the DMSFE baseline, and the DMSFE tilted densities. Let $\widehat{W}_{h,t}$ and $\widehat{W}_{g,t}$ be the corresponding realized wealth at time t , where $\widehat{W}_{h,t}$ and $\widehat{W}_{g,t}$ are functions of time t realized excess return, r_t , as well as the optimal allocations to the risky and risk-free assets computed in (24) and (25)

$$\begin{aligned} \widehat{W}_{h,t} &= (1 - \hat{\omega}_{h,t-1}) \exp(r_{f,t-1}) + \hat{\omega}_{h,t-1} \exp(r_{f,t-1} + r_t) \\ \widehat{W}_{g,t} &= (1 - \hat{\omega}_{g,t-1}) \exp(r_{f,t-1}) + \hat{\omega}_{g,t-1} \exp(r_{f,t-1} + r_t). \end{aligned} \quad (26)$$

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 out-of-sample period as follows

$$CER_k = \left((1-A)(T-m)^{-1} \sum_{\tau=m+1}^T \widehat{U}_{k,\tau} \right)^{12/(1-A)} - 1, \quad (27)$$

where $k \in \{base, tilt, bench\}$ and $\widehat{U}_{\tau,k} = \widehat{W}_{k,\tau}^{1-A}/(1-A)$ denotes the time- τ realized utility associated with predictive density k . Finally, we compute the certainty equivalent return

differential between models g and h as

$$CERD_{g,h} = CER_g - CER_h, \quad g \in \{base, tilt\} \quad (28)$$

[Table 3](#) reports the annualized CERD estimates associated with the DMSFE tilted and baseline density forecasts for the equity premium, relative to the HA. We report CERD values for the entire out-of-sample period, 1995–2010, as well as for the three sub-samples 1995–2000, 2001–2005, and 2006–2010. We also examine the sensitivity of the results to different risk preferences by considering a range of values for the risk-aversion coefficient, $A = 2, 3, 5, 10$.¹²

Overall, the tilted DMSFE predictive densities produce higher risk-adjusted returns than their HA counterparts. During 1995–2010, their implied annualized percent CERDs lie between 1% and 5.2%, depending on the value of the risk aversion coefficient. The corresponding CERDs for the baseline predictive densities range between -0.4% and -0.07%. Switching to the different sub-samples, we find that between 1995–2000, the CERDs based on the tilted densities are between 2.6% and 14.8%, while their baseline counterparts lie between -3.7% and -0.7%. As for the two more recent sub-samples, 2001–2005 and 2006–2010, we see that the magnitudes of the CERDs for the tilted densities are much smaller with values between 0.11% and 1% depending on the sub-sample and risk aversion coefficient. Although the tilted CERDs are smaller than the baseline CERDs during 2001–2005, they are higher than their baseline counterparts between 2006 and 2010 for all levels of risk aversion.

¹²We compute the optimal portfolio weights for the CRRA investor using the approximation proposed by [Campbell and Viceira \(2001\)](#), without imposing any constraints to preclude short-selling or leverage.

5 Conclusions

The paper introduces a novel approach to improve density forecasts for the equity premium using information from the derivatives markets in a time-series setting. We tilt predictive densities from a state-of-the-art model in the empirical asset-pricing literature towards the second moment of the distribution implied by option prices, after removing the variance risk premium using a simple regression-based approach. 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 U.S. equity premium based on the model combination approach of [Rapach and Zhou \(2012\)](#), we show that tilting significantly improves both the statistical and economic predictability of stock returns, based on standard metrics from the density forecast comparison and evaluation literature. 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 methodology to higher moments, such as skewness and kurtosis, which have been receiving increased attention in empirical asset pricing, is a research agenda worth pursuing, especially as more options data become available.

References

- ALTAVILLA, C., R. GIACOMINI, AND R. CONSTANTINI. 2014. “BOND RETURNS AND MARKET EXPECTATIONS.” *Journal of Financial Econometrics*, 12(4): 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.
- BACK, K. 2010. *Asset Pricing and Portfolio Choice Theory.*: OXFORD UNIVERSITY PRESS.
- BARBERIS, N. 2000. “INVESTING FOR THE LONG RUN WHEN RETURNS ARE PREDICTABLE.” *Journal of Finance*, LV 225–264.
- BERKOWITZ, J. 2001. “TESTING DENSITY FORECASTS, WITH APPLICATIONS TO RISK MANAGEMENT.” *Journal of Business & Economic Statistics*, 19 465–474.
- BLOOM, N. 2009. “THE IMPACT OF UNCERTAINTY SHOCKS.” *Econometrica*, 77(3): 623–685.
- BOLLERSLEV, T., G. TAUCHEN, AND H. ZHOU. 2009. “EXPECTED STOCK RETURNS AND VARIANCE RISK PREMIA.” *Review of Financial Studies*, 22(11): 4463–4492.
- CAMPBELL, J., AND L. VICEIRA. 2001. “WHY SHOULD WE BUY LONG-TERM BONDS?” *American Economic Review*, 91 99–127.
- CENESIZOGLU, T., AND A. TIMMERMANN. 2012. “DO RETURN PREDICTION MODELS ADD ECONOMIC VALUE?” *Journal of Banking and Finance*, 36 2974–2987.

- CHRISTOFFERSEN, P., K. JACOBS, AND B.Y. CHANG. 2013. “FORECASTING WITH OPTION IMPLIED INFORMATION.” IN *Handbook of Economic Forecasting Volume 2A*: ELSEVIER.
- CLARK, T., 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.
- DOORNIK, J.A., AND H. HANSEN. 2008. “AN OMNIBUS TEST FOR UNIVARIATE AND MULTIVARIATE NORMALITY.” *Oxford Bulletin of Economics and Statistics*, 70 927–939.
- GIACOMINI, R., AND G. RAGUSA. 2013. “THEORY-COHERENT FORECASTING.” TECHNICAL REPORT, UNIVERSITY COLLEGE LONDON AND LUISS UNIVERSITY.
- GOYAL, A., AND I. WELCH. 2008. “A COMPREHENSIVE LOOK AT THE EMPIRICAL PERFORMANCE OF EQUITY PREMIUM PREDICTION.” *Review of Financial Studies*, 21(4): 1455–1508.
- JORION, P. 1995. “PREDICTING VOLATILITY IN THE FOREIGN EXCHANGE MARKET.” *Journal of Finance*, 50(2): 507–528.

- KANDEL, S., AND R. STAMBAUGH. 1996. "ON THE PREDICTABILITY OF STOCK RETURNS: AN ASSET-ALLOCATION PERSPECTIVE." *Journal of Finance*, 51 385–424.
- KRÜGER, F., T. CLARK, AND F. RAVAZZOLO. 2015. "USING ENTROPIC TILTING TO COMBINE BVAR FORECASTS WITH EXTERNAL NEWCASTS." *Journal of Business & Economic Statistics*, FORTHCOMING.
- LJUNG, G.M., AND G.E.P. BOX. 1978. "ON A MEASURE OF LACK OF FIT IN TIME SERIES MODELS." *Biometrika*, 65 297–303.
- POON, S-H., AND C.W.J GRANGER. 2003. "FORECASTING VOLATILITY IN FINANCIAL MARKETS: A REVIEW." *Journal of Economic Literature*, 41(2): 478–539.
- RAPACH, D.E., J.K. STRAUSS, AND G. ZHOU. 2010. "OUT-OF-SAMPLE EQUITY PREMIUM PREDICTION: COMBINATION FORECASTS AND LINKS TO THE REAL ECONOMY." *Review of Financial Studies*, 23 821–862.
- RAPACH, D.E., AND G. ZHOU. 2012. "FORECASTING STOCK RETURNS." IN *Handbook of Economic Forecasting Volume 2A*: ELSEVIER.
- 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.
- ROSSI, B., AND T. SEKHPOSYAN. 2014. "EVALUATING PREDICTIVE DENSITIES OF U.S. OUTPUT GROWTH AND INFLATION IN A LARGE MACROECONOMIC DATA SET." *International Journal of Forecasting*, 30 662–682.
- 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.

Tables

Table 1: Density calibration diagnostics

Test	Tilted DMSFE	Baseline DMSFE	HA
Kolmogorov-Smirnov	0.063	0.049	0.085
Anderson-Darling	0.067	0.009	0.007
Ljung-Box I	0.447	0.657	0.645
Ljung-Box II	0.722	<0.001	<0.001
Berkowitz I	0.875	<0.001	<0.001
Berkowitz II	0.299	0.146	0.130
Berkowitz III	0.703	0.001	<0.001
Doornik-Hansen	<0.001	<0.001	<0.001

Note: We report p-values for all tests. We use DMSFE to refer to discounted mean squared forecast error densities and HA to refer to historical average densities. The Kolmogorov-Smirnov and the Anderson-Darling are uniformity tests for the probability integral transformation (PIT). The Ljung-Box I and Ljung-Box II tests are independence tests for the 1st and 2nd central PIT moments. The Berkowitz tests are based on the normal inverse transformation of the PITs and pertain to normality (I), independence (II), and both normality & independence (III). The Doornik-Hansen is a normality test for the the normal inverse transformation of the PITs based on the squares of skewness and kurtosis measures. See [Section 4.2](#) for details.

Table 2: Forecast performance: average log score differentials

	Tilted DMSFE-HA	Baseline DMSFE-HA
1995-2000	0.0969	-0.0037
2001-2005	0.2121	0.0026
2006-2010	0.2255	0.0032
Full Sample	<0.001	0.436

Note: We use DMSFE to refer to discounted mean squared forecast error densities and HA to refer to historical average densities. We report average monthly log score differences for the tilted DMSFE-HA and baseline DMSFE-HA pairs of density forecasts for three different periods in our sample: (i) 1995–2000, (ii) 2001–2005, and (iii) 2006–2010. The last row of the table reports the p-values for a Diebold-Mariano test based on the log score differences for the tilted DMSFE-HA and baseline DMSFE-HA pairs over the entire out-of-sample period. See [Section 4.2](#) for details.

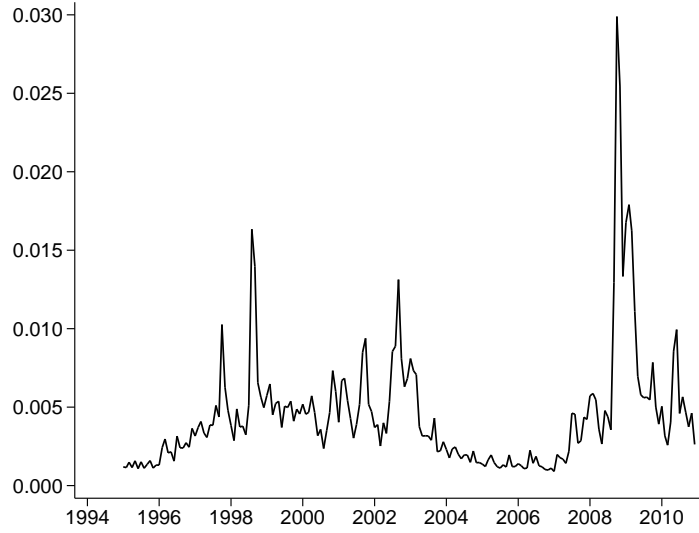
Table 3: Economic performance

	Tilted DMSFE-HA				Baseline DMSFE-HA			
	A=2	A=3	A=5	A=10	A=2	A=3	A=5	A=10
1995-2000	14.82%	9.31%	5.30%	2.55%	-3.66%	-2.36%	-1.42%	-0.66%
2001-2005	0.95%	0.62%	0.41%	0.23%	2.09%	1.39%	0.84%	0.46%
2006-2010	0.40%	0.38%	0.23%	0.11%	0.27%	0.26%	0.16%	0.04%
Full Sample	5.18%	3.41%	2.02%	0.99%	-0.44%	-0.27%	-0.17%	-0.07%

Note: We use DMSFE to refer to discounted mean squared forecast error densities and HA to refer to historical average densities. We report the relative return performance (percent, annualized) of portfolios for the tilted and baseline DMSFE density forecasts versus those based on the HA density forecasts for the full sample and three different periods: (i) 1995–2000, (ii) 2001–2005, and (iii) 2006–2010. A positive number indicates that the DMSFE density forecasts generate higher risk-adjusted returns than the historical average. We examine the sensitivity to risk preferences by considering alternative values of the coefficient of relative risk aversion A . See [Section 4.3](#) for details.

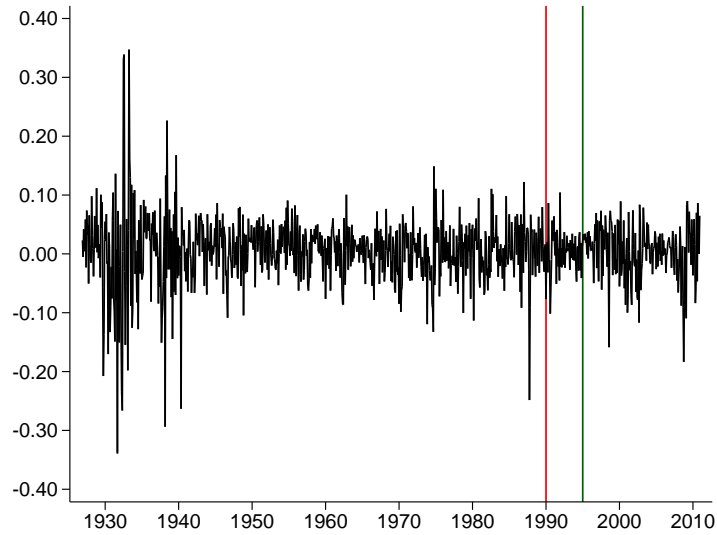
Figures

Figure 1: Monthly risk-neutral variance



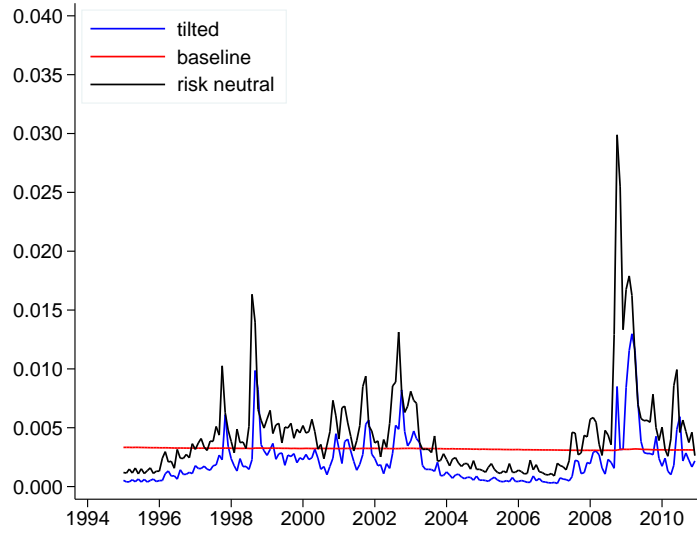
Note: We plot the end of month values of the CBOE VIX squared as in [Bollerslev et al. \(2009\)](#).

Figure 2: Monthly U.S. equity premium

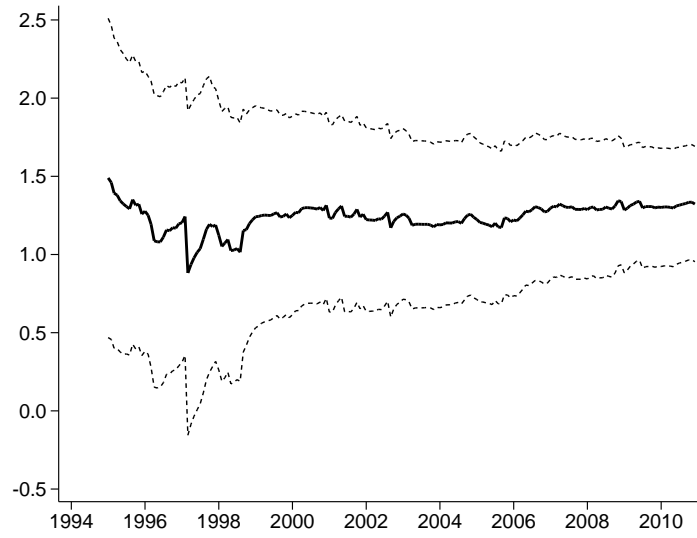


Note: We plot the US monthly equity premium between January 1926 and December 2010 using end-of-month returns for the S&P 500 and the risk-free rate as in [Goyal and Welch \(2008\)](#). The red vertical line indicates January 1990, when VIX becomes available, and the green vertical line indicates January 1995, the first date when we tilt the baseline predictive densities.

Figure 3: Physical and risk-neutral variance series



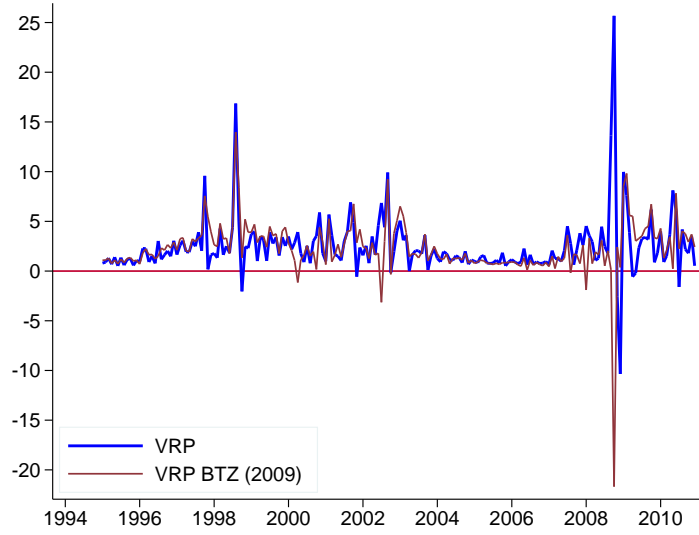
(a) Variance



(b) Variance risk premium regression slope

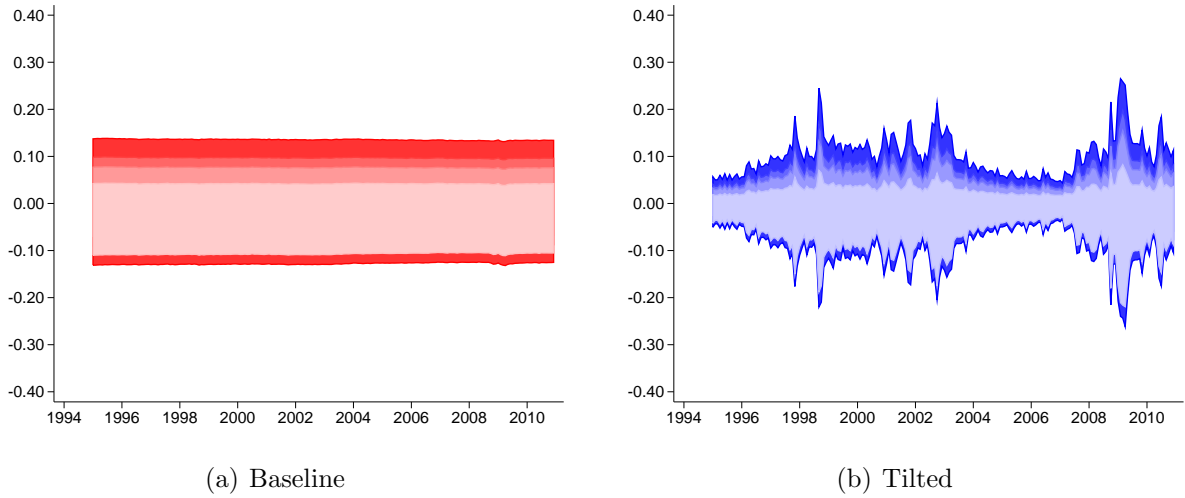
Note: Panel (a) of the figure contains monthly time series of the variance for the tilted (blue), baseline (red), and risk-neutral (black) distributions. We use the end-of-month values of the squared VIX as our measure of risk-neutral variance. Panel (b) of the figure contains the estimate and associated 95% confidence intervals for the slope coefficient in the variance risk premium regression described in [Section 2.3](#) using an expanding window with initial size of 60 monthly observations starting in January 1995.

Figure 4: Variance risk premium



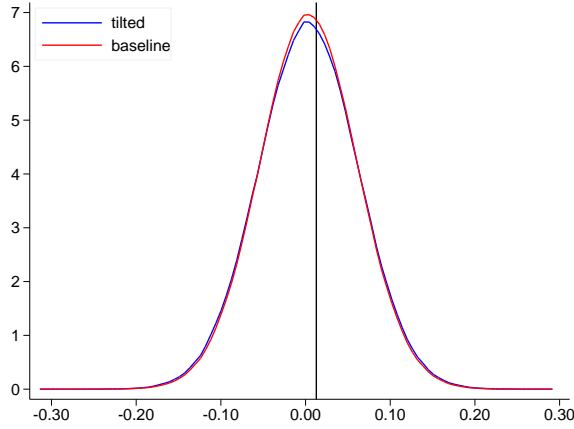
Note: We plot the our variance risk premium (VRP) along with the variance risk premium from [Bollerslev et al. \(2009\)](#).

Figure 5: Posterior probability intervals

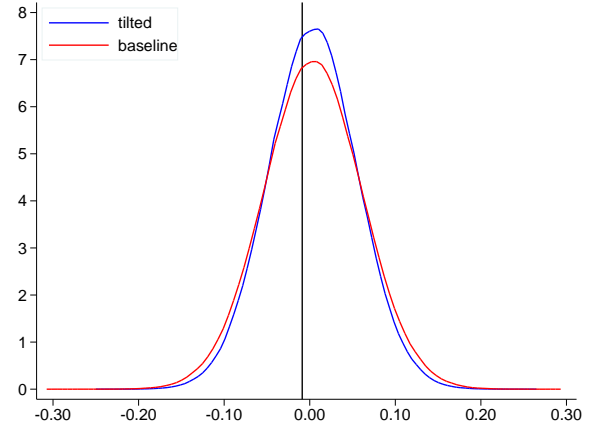


Note: We use DMSFE to refer to discounted mean squared forecast error densities. We plot the posterior probability intervals for the baseline and tilted DMSFE densities over the out-of-sample period. The shading of varying intensity indicates the 1%–99%, 5%–95%, 10%–90% and 25%–75% interquantile ranges for the DMSFE density forecasts in each month.

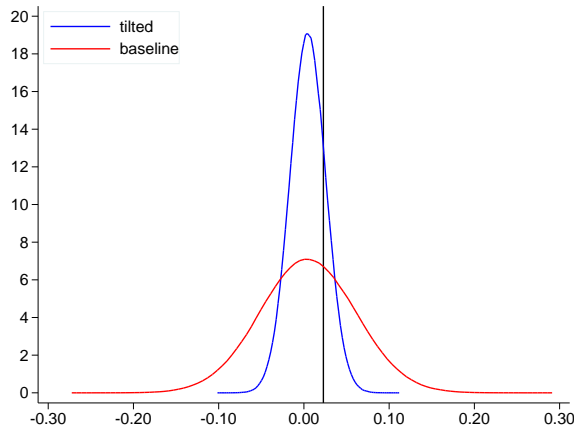
Figure 6: Illustrating the effect of tilting



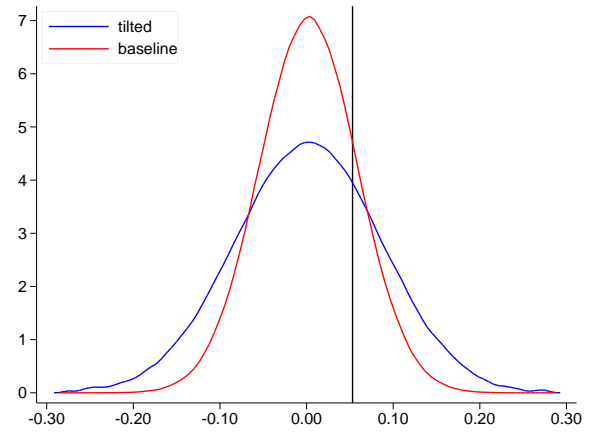
(a) 1997:12, KLIC=0.0007



(b) 1999:08, KLIC=0.0089



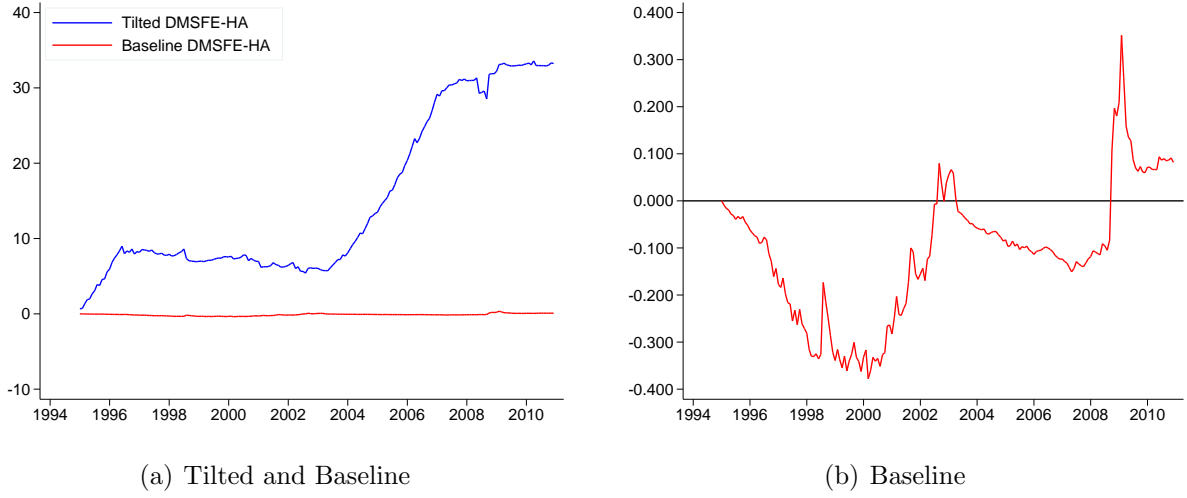
(c) 2006:01, KLIC=0.5588



(d) 2009:05, KLIC=0.2162

Note: We use DMSFE to refer to discounted mean squared forecast error densities. We provide kernel density plots for the baseline and tilted DMSFE predictive densities for 4 distinct forecast dates within the out-of-sample period. The black vertical line indicates the observed value of the equity premium for the forecast date under consideration.

Figure 7: Cumulative log score differences



Note: We use DMSFE to refer to discounted mean squared forecast error densities and HA to refer to historical average densities. Using $LS_{base}(t)$, $LS_{tilt}(t)$, and $LS_{HA}(t)$ to denote the log scores associated with the baseline DMSFE, tilted DMSFE, and HA predictive densities for month t , we plot the cumulative sums of the following log score differences indicated by the red and blue lines, respectively: $CLSD(t)_{base,HA} = \sum_{\tau=m+1}^t LS_{base}(\tau) - LS_{HA}(\tau)$ and $CLSD(t)_{tilt,HA} = \sum_{\tau=m+1}^t LS_{tilt}(\tau) - LS_{HA}(\tau)$, where $t = m + 1, \dots, T$. For each month t , we estimate the forecasting models recursively and generate one-month ahead density forecasts, which are used to calculate the cumulative log-score differences. Due to large scale differences for $CLSD(t)_{tilt,HA}$ and $CLSD(t)_{base,HA}$ in panel (a), we plot $CLSD(t)_{base,HA}$ in panel (b) using a smaller scale.