Do Return Prediction Models Add Economic Value?[*]

Tolga Cenesizoglu
HEC Montréal and CIRPÉE

Allan Timmermann
UCSD and CREATES

May 25, 2011

Abstract

Return prediction models with a time-varying mean often fail to improve on the out-of-sample mean squared error of a constant equity premium model. However, this does not rule out that return prediction models that allow for a time-varying probability distribution can add economic value by helping improve investors' portfolio choice. This paper shows that statistical and economic measures of forecasting performance weight forecast errors very differently and that return forecasts from models with time-varying mean and variance, when used to guide the portfolio choice of an investor with power utility, can lead to significant improvements over the forecasts from a model that assumes a constant return distribution. Specifically, models with constant mean and volatility tend to overestimate the right tail of the return distribution and so lead to stock allocations that are on average too large with resulting lower average utility for risk averse investors. Our results demonstrate that return prediction models can add economic value even when they fail to produce accurate forecasts of mean returns and suggest the need for focusing on broader measures of distributional accuracy when evaluating the economic value of return prediction models.

Key words: predictability of stock returns, mean squared forecast error, portfolio selection, probability distribution forecasts.

---

[*] Seminar participants at HEC, University of Montreal, Imperial College, London, University of Toronto, Goldman Sachs and CREATES, University of Aarhus, are thanked for helpful comments on an earlier version of this paper. Timmermann acknowledges support from CREATES, funded by the Danish National Research Foundation.

1 Introduction

The recent debate on predictability of stock market returns has focused on the ability of models with a time-varying mean to generate better out-of-sample return forecasts than a simple model that assumes constant expected returns (see, e.g., Lettau and Ludvigsson (2001), Ang and Bekaert (2007), Campbell and Thompson (2008), Welch and Goyal (2008), and Rapach, Strauss and Zhou (2010)). Forecast precision has typically been measured by out-of-sample mean squared forecast errors or, equivalently, the out-of-sample R^2. A clear limitation of such measures is that they capture statistical forecasting performance and thus need not reveal the economic value of return forecasts when used to guide portfolio decisions. It is possible for a return prediction model to perform poorly in statistical terms when compared to a constant equity premium model, yet at the same time add significant economic value when its forecasts are used for portfolio choice decisions. This situation can arise because forecast errors of different magnitudes and different signs (i.e., over- and under-predictions of returns) are weighted very differently by economic and statistical loss functions.

This paper shows that the distinction between statistical and economic measures of forecasting performance is very important empirically. Comparing the empirical performance of several return prediction models, our analysis finds only a weak relation between statistical measures of predictive accuracy such as mean squared forecast errors and economic measures such as the certainty equivalent return based on using return forecasts to guide the portfolio selection of a risk averse investor. In contrast, there is a significant relation between the accuracy of forecasts of the return probability distribution and the economic value of such forecasts. The reason for this finding is clear: Risk averse investors require an estimate of the full probability distribution of future stock returns to make their portfolio decisions and so it is necessary to go beyond expected returns−emphasized by forecasts that minimize mean squared forecast errors−and consider the accuracy with which the full probability distribution is predicted.

When it comes to the economic cost of forecast errors, not all errors are created equal and so it is important to carefully consider predictability of different parts of the return distribution. Predictability of the tails of the return distribution turns out to be particularly important for understanding the economic value of return forecasts. Forecasts from constant mean and volatility models significantly overestimate the probability of outcomes in the right tail of the return distribution and conversely underestimate the probability of outcomes in the center. This leads to an overly risky asset allocation and to too frequent periods with large negative wealth shocks which are very costly to risk averse investors. Models that allow for a time-varying mean and volatility produce better density forecasts, more conservative portfolio choices and higher average utility.

Our study analyzes a variety of return prediction models that differ in whether they assume predictability in the mean and/or variance of returns and in how much structure they impose on the return distribution. As a simple no-predictability benchmark in the spirit of Welch and Goyal (2008), we consider both a prevailing (constant) mean and variance model with normally distributed returns as well as a prevailing quantile model that assumes a constant return distribution but does not impose normality. We then consider time-varying mean, time-varying volatility and time-varying mean and volatility specifications in the EGARCH class studied by Glosten, Jagannathan and Runkle (1993). We further consider a flexible approach that models predictability at individual points (quantiles) of the return distribution and so generalizes existing work on predictability of the mean and variance of returns. By focusing separately on quantiles in different parts of the return distribution, e.g., the left and right tails, we can directly address whether those parts of the return distribution are predictable.

The main contributions of our paper are as follows. Our first contribution is to analyze where commonly used return prediction models succeed and where they fail when it comes to predicting the probability distribution of returns. Surprisingly little is known about which parts of the return distribution are predictable and how they depend on economic state variables. To address this point, we introduce a range of probability distribution tests that diagnose whether distribution forecasts are correctly specified, which parts of the return distribution (e.g., the tails or center) cause predictive failure and whether one model produces better distribution forecasts than another.

Consistent with the evidence in Welch and Goyal (2008), we find that the simple constant mean model generates lower out-of-sample root mean squared forecast errors than more complicated models that require the estimation of additional parameters. However, our analysis also shows that this model is clearly misspecified and produces relatively poor forecasts of the center and right tail of the return distribution. Moreover, when this model is used to generate distribution forecasts, it comes up short relative to models that consider dynamics in both the mean and volatility of stock returns. The best models when measured by the precision of their out-of-sample distribution forecasts are EGARCH specifications that allow for a time-varying mean and volatility. Dynamic quantile models tend to produce good distribution forecasts in-sample, but poor out-of-sample forecasts due to the difficulty of precisely estimating the parameters of the tail quantiles.

Our second contribution is to evaluate the economic significance of predictability of the distribution of stock returns. We do so in two ways. First, we use our forecasts of the return distribution in an out-of-sample asset allocation exercise that combines stocks and T-bills for investors with power utility. Second, we use the non-parametric stochastic discount factor approach advocated by Almeida and Garcia (2008) to compute model-free estimates of the risk-adjusted returns associated with the different prediction models and conduct pair-wise model comparisons.

Our empirical estimates of the economic performance measures suggest that allowing for a time-varying mean and volatility leads to significantly improved portfolio performance, whereas time-variations in individual quantiles is more difficult to capture out-of-sample. Specifically, we find that EGARCH models with a time-varying mean and volatility generate certainty equivalent returns that are 150 basis points per annum higher than those of the model that assumes a constant return distribution. Moreover, the risk-adjusted return performance of these models exceeds that of the constant distribution model by 200 basis points per annum.

Our third contribution is to show how statistical and economic measures of forecasting performance are related. We conduct an empirical rank correlation analysis of the relation between two statistical measures of forecasting performance tracking mean squared forecast errors and the accuracy of probability distribution forecasts versus economic measures of forecasting performance such as certainty equivalent returns or the Sharpe ratio. The measure of predictive accuracy based on the entire probability distribution of returns closely tracks the economic value of return predictions. In contrast, mean squared forecast error performance is only weakly related to the models' economic value.

Our fourth and final contribution is to evaluate and quantify the importance of parameter estimation error. Many studies have emphasized the importance of considering out-of-sample as opposed to in-sample forecasting performance. Out-of-sample forecasting performance addresses whether an investor in real time could have made use of a forecasting model to generate economically valuable return forecasts. It accounts for the effect of parameter estimation error which tends to reduce the precision of return forecasts and so leads to worse performance. By comparing in-sample and out-of-sample certainty equivalent return values we can evaluate the importance of parameter estimation error. Our results suggest that parameter estimation error leads to a reduction in certainty equivalent returns ranging from 80 to 210 basis points per year. Economic loss from estimation error is largest for the dynamic quantile models and smallest for the EGARCH models.

While the focus of our analysis is on predictability of the full probability distribution of returns and its economic significance, our paper is clearly related to the large literature on predictability of the mean and volatility of stock returns. Empirical results have been somewhat mixed, indicating that while full-sample predictability of mean returns is quite strong, out-of-sample predictability of mean returns is considerably weaker and can be difficult to exploit. Ang and Bekaert (2007), Campbell (1987), Campbell and Shiller (1988), Campbell and Thompson (2008), Cochrane (2008), Fama and French (1988, 1989), Ferson (1990), Ferson and Harvey (1993), Lettau and Ludvigsson (2001) and Pesaran and Timmermann (1995) find some evidence of predictability of the mean. However, this evidence has been questioned by Bossaerts and Hillion (1999) and Welch and Goyal (2008) who argue that the parameters of return prediction models are estimated with insufficient precision to render *ex-ante* return forecasts of much value.[1] Similarly, while the volatility of stock returns is known to follow a pronounced counter-cyclical pattern (Schwert (1989)), there is relatively weak evidence that macroeconomic state variables contain information useful for predicting stock market volatility. Engle, Ghysels and Sohn (2007) find some evidence that inflation volatility helps predict the volatility of stock returns. However, the volatility of interest rate spreads and growth in industrial production, GDP or the monetary base fail to consistently predict future volatility, with evidence being particularly weak in the post-WWII sample. This is consistent with findings in Paye (2010) and Ghysels, Santa-Clara and Valkanov (2006).

The outline of the paper is as follows. Section 2 introduces the models used to capture predictability of the return distribution, presents the data and reports empirical results. Section 3 evaluates the statistical forecasting performance of the various models, while Section 4 considers the economic value of return forecasts in portfolio selection experiments and relates the statistical and economic measures of forecasting performance. Finally, Section 5 concludes.

2 Models of Predictability in the Distribution of Stock Returns

The literature on return predictability is very extensive but has largely focused on predictability of the mean and variance of stock returns. Our analysis broadens the perspective by considering predictability in the full probability distribution of stock returns and asking which (if any) parts of the return distribution are predictable. The aim of our analysis is not to consider an exhaustive list of modeling approaches or predictor variables for capturing time-variations in the return distribution. Rather, we focus on models in common use such as linear regressions and time-varying volatility models and use four common predictor variables. Quantile models are added to the mix in order to study the effect of making less restrictive distributional assumptions and to focus on return predictability at separate points of the return distribution.

---

[1] Dangl and Halling (2008), Johannes, Korteweg and Polson (2009) and Paye and Timmermann (2006) attribute the poor out-of-sample forecasting performance to parameter instability, while Lettau and van Nieuwerburgh (2008) propose shifts in the mean of predictor variables such as the dividend yield as a source of predictive failure.

As a simple no-predictability benchmark, consider a constant mean and volatility model:

$$r_{t+1} = \beta_0 + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim N(0, \sigma^2). \tag{1}$$

Here r_{t+1} is the stock return computed in excess of a safe T-bill rate. This model is equivalent to the "prevailing mean" (PM) model of Welch and Goyal (2008) with the addition of a constant volatility and the assumption of normally distributed returns. Later we relax the assumption of normally distributed returns, but maintain the constant mean and volatility assumptions.

2.1 Time-variations in the Mean and Volatility of Returns

To explore whether allowing for time-variations in the mean, volatility, or both, improves return forecasts and to gauge the effect of parameter estimation error, predictability of the mean and volatility of returns is analyzed in separate steps.

The effect of allowing for a time-varying mean (TVM), but holding volatility constant, is considered through the following model:

$$r_{t+1} = \beta_0 + \beta_1' x_t + \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim N(0, \sigma^2), \tag{2}$$

where x_t is a set of state variables known at time t. Conversely, a model that allows for a time-varying volatility, but holds the mean constant can be studied through an EGARCH specification of the type proposed by Glosten, Jagannathan, and Runkle (1993):

$$\begin{aligned} r_{t+1} &= \beta_0 + \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim N(0, \sigma_{t+1}^2), \\ \log(\sigma_{t+1}^2) &= \delta_0 + \delta_2 \log(\sigma_t^2) + \delta_3 \left| \frac{\varepsilon_t}{\sigma_t} \right| + \delta_4 \frac{\varepsilon_t}{\sigma_t}. \end{aligned} \tag{3}$$

In this model the conditional volatility, σ_{t+1}, depends on the lagged volatility, σ_t, as well as the return innovation, ε_t, whose effect can depend on whether it is positive or negative. This model is naturally extended to a time-varying mean and volatility EGARCH specification:

$$\begin{aligned} r_{t+1} &= \beta_0 + \beta_1' x_t + \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim N(0, \sigma_{t+1}^2), \\ \log(\sigma_{t+1}^2) &= \delta_0 + \delta_2 \log(\sigma_t^2) + \delta_3 \left| \frac{\varepsilon_t}{\sigma_t} \right| + \delta_4 \frac{\varepsilon_t}{\sigma_t}. \end{aligned} \tag{4}$$

We refer to this as the TVM-EGARCH model.

Finally, the most general model, labeled the time-varying mean EGARCHX (TVM-EGARCHX) model, allows the economic state variables to affect both the mean and volatility equations:

$$\begin{aligned} r_{t+1} &= \beta_0 + \beta_1' x_t + \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim N(0, \sigma_{t+1}^2), \\ \log(\sigma_{t+1}^2) &= \delta_0 + \delta_1' x_t + \delta_2 \log(\sigma_t^2) + \delta_3 \left| \frac{\varepsilon_t}{\sigma_t} \right| + \delta_4 \frac{\varepsilon_t}{\sigma_t}. \end{aligned} \tag{5}$$

The models in equations (**??**)-(**??**) are all nested by Eq. (**??**). However, for purposes of forecasting the return distribution out-of-sample it is possible that the simpler models dominate since they have fewer parameters to estimate. Such EGARCH models smooth the effect of state variables over different parts of the return distribution. If the effect of predictor variables varies across different parts of the return distribution, this can lead to misspecified distribution models. To deal with this issue, we next consider a flexible approach that models individual quantiles of the return distribution.

2.2 Quantile Models

To understand how different parts of the return distribution vary over time, it is helpful to consider individual quantiles located at separate points of the return distribution. Let $\alpha \in (0, 1)$ represent a particular quantile of interest. By varying α from values near zero (representing draws from the left tail of the return distribution) through one-half (representing the center) to values near one (representing the right tail), variations in the complete return distribution can be tracked. Joint consideration of such quantiles provides a rich picture of variations in the return distribution.

The prevailing (constant) quantile (PQ) model assumes a constant return distribution−but does not impose normally distributed returns and thus takes the form

$$q_\alpha(r_{t+1}|\mathcal{F}_t) = \beta_{0,\alpha}, \tag{6}$$

where $q_\alpha(r_{t+1}|\mathcal{F}_t)$ is the conditional $\alpha-$quantile given current information, $\mathcal{F}_t$. To see if dynamics in past quantiles and lagged returns help improve predictions of the individual quantiles, a simple dynamic quantile (DQ) specification along the lines of Engle and Manganelli (2004) is also considered:

$$q_\alpha(r_{t+1}|\mathcal{F}_t) = \beta_{0,\alpha} + \beta_{2,\alpha} q_\alpha(r_t|\mathcal{F}_{t-1}) + \beta_{3,\alpha}|r_t|, \tag{7}$$

where $|r_t|$ is the absolute return. Finally, a general quantile specification that includes economic covariates, x_t, and also allows for dynamic effects from past quantiles and lagged returns is considered:

$$q_\alpha(r_{t+1}|\mathcal{F}_t) = \beta_{0,\alpha} + \beta'_{1,\alpha} x_t + \beta_{2,\alpha} q_\alpha(r_t|\mathcal{F}_{t-1}) + \beta_{3,\alpha}|r_t|. \tag{8}$$

This specification−labeled the DQX model−is consistent with volatility clustering and persistent tail dynamics in stock returns.[2] The local effect of x_t on individual $\alpha-$quantiles is assumed to

---

[2] Foresi and Peracchi (1995) characterize the cumulative distribution function of stock returns as a function of a set of economic state variables. In effect they model the "dual" of the quantile function and estimate conditional logit models over a grid of values for the cumulative distribution function of returns.

be linear but the model is very flexible since the slope coefficient ($\beta_{1,\alpha}$) is allowed to differ across quantiles.

To gain intuition for the quantile models, note that the largest impact of economic predictor variables is expected in the tails of the return distribution if the effect of such variables arises through a volatility risk premium. To see this, suppose that return volatility varies in proportion with the state variables, x_t, and that it earns a risk premium, κ (see, e.g., Merton (1980)):

$$\begin{aligned} r_{t+1} &= \mu + \kappa\sigma_{t+1} + \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim N(0, \sigma_{t+1}) \\ \sigma_{t+1} &= \varphi_0 + \varphi_1' x_t, \end{aligned} \tag{9}$$

where φ_1 measures the volatility effect of x_t. This specification implies conditional quantiles of the form

$$q_\alpha(r_{t+1}|\mathcal{F}_t) = \mu + \varphi_0(\kappa + q_\alpha^N) + (\kappa + q_\alpha^N)\varphi_1' x_t \equiv \beta_{0,\alpha} + \beta_{1,\alpha}' x_t, \tag{10}$$

where the slope coefficient is $\beta_{1,\alpha} = (\kappa + q_\alpha^N)\varphi_1$ and q_α^N is the $\alpha-$quantile of the normal distribution which takes on larger (absolute) values further out in the tails and shifts sign from negative to positive as α moves from values below the median to values above it. Economic theory suggests that $\kappa > 0$, so variables that are positively correlated with volatility ($\varphi_1 > 0$) can be expected to have negative slope coefficients in the quantile regression sufficiently far in the left tail (small $\alpha-$values) and positive coefficients above the median. The reverse pattern arises for variables correlating negatively with volatility ($\varphi_1 < 0$).

2.3 Estimation

Estimation of EGARCH models such as those in Eq. (**??**) can be based on maximum likelihood methods, whereas estimation of the parameters of the quantile models is more non-standard. Following Koenker and Bassett (1978), quantiles are estimated by replacing the conventional quadratic loss function underlying most empirical work on return predictability with a "tick" loss function

$$L_\alpha(e_{t+1}) = (\alpha - \mathbf{1}\{e_{t+1} < 0\})e_{t+1}, \tag{11}$$

where $e_{t+1} = r_{t+1} - \hat{q}_{\alpha,t}$ is the forecast error, $\hat{q}_{\alpha,t} = q_\alpha(r_{t+1}|\mathcal{F}_t)$ is short-hand notation for the conditional quantile forecast computed at time t and $\mathbf{1}\{\cdot\}$ is the indicator function. Under this objective function, the optimal forecast is the conditional quantile. To obtain estimates of the parameters of the dynamic quantile specification in Eq. (**??**), we adopt the tick-exponential quasi maximum likelihood estimation approach of Komunjer (2005). Estimates of the parameters $\theta_\alpha =$

$(\beta_{0,\alpha}, \beta_{1,\alpha}, \beta_{2,\alpha}, \beta_{3,\alpha})$ solve the objective

$$\hat{\theta}_\alpha = \arg\max_{\theta_\alpha} \left\{ T^{-1} \sum_{t=1}^{T} \ln \varphi_t^\alpha(r_t, q_\alpha(r_t|\mathcal{F}_{t-1}, \theta_\alpha)) \right\}, \tag{12}$$

where φ_t^α is a probability density from the tick-exponential family:

$$\varphi_t^\alpha(r_t, q_\alpha) = \exp(-\frac{1}{\alpha}(q_\alpha - r_t)\mathbf{1}\{r_t \leq q_\alpha\} + \frac{1}{1-\alpha}(q_\alpha - r_t)\mathbf{1}\{r_t > q_\alpha\}). \tag{13}$$

Komunjer (2005) establishes conditions under which the parameter estimates, $\hat{\theta}_\alpha$, are asymptotically normally distributed and provides methods for estimating their standard errors.

2.4 Data

Our empirical analysis uses a data set comprising monthly stock market returns along with a set of four predictor variables that have been extensively used in finance, namely (i) the dividend yield, measured by dividends during the previous twelve months divided by the value of the market index; (ii) the 3-month T-bill rate; (iii) the term spread, measured by the difference between the yield on long-term government bonds and the three-month T-bill rate; and (iv) the default yield measured by the spread between BAA and AAA rated corporate bonds. Our analysis is limited to these four predictor variables due to their widespread use in studies on return predictability. The dividend yield has been a workhorse in this literature (e.g., Fama and French (1988), and Campbell and Shiller (1988)) as has the T-bill rate (Campbell (1987), and Ang and Bekaert (2007)). Similarly, the term spread and default yield have been used extensively to track a time-varying risk component in expected returns (Fama and French (1988)).

Stock market returns are continuously compounded and are measured by the S&P500 index including dividends. A short T-bill rate is subtracted to obtain excess returns. The data sample runs from 1926:01 through 2008:12 for all variables. Such a long sample period is important in order to evaluate predictability in the tails of the return distribution.

2.5 Empirical Estimates

Estimates of the slope coefficients on the economic state variables in the TVM (linear regression), TVM-EGARCH and TVM-EGARCHX models are presented in Table 1. Only the dividend yield has a significant effect on the conditional mean in all three models, while the T-bill rate is also significant in one of the EGARCH models. In contrast, all state variables are significant for the conditional volatility in the TVM-EGARCHX model. Thus the economic state variables only have weak in-sample predictive power over the mean, but do have considerable predictive power over the volatility of monthly stock returns. Lagged volatility and past return innovations also help predict future volatility, with positive shocks having a smaller impact on volatility than negative shocks of the same magnitude.

Turning to the quantile regressions, Table 2 reports estimates of the slope coefficients ($\beta_{1,\alpha}$) for each of the predictor variables. Quantiles in the range $\alpha \in \{0.05, 0.10, 0.20, ..., 0.90, 0.95\}$ are considered as quantiles further out in the tails than 0.05 and 0.95 are not as precisely estimated. Some of the predictor variables are highly persistent and so are correlated with the lagged quantiles. To make the results easier to interpret, the table reports estimates from a quantile model that excludes dynamic effects from past quantiles.

The standard linear return prediction model assumes that economic state variables have the same effect on the return distribution across all quantiles so $\beta_{1,\alpha} = \beta_1$ for all values of α.[3] This is not supported by the results in Table 2. In fact, the slope coefficients for the dividend yield and the term spread are only systematically significant for the left-tail quantiles, $\alpha = 0.05, 0.10, .., 0.30$. In both cases the coefficient estimates are positive, suggesting that narrower term spreads (typically occurring during recessions) or lower dividend yields are associated with thicker left tails, indicating higher risk. Conversely, the coefficients on the T-bill rate are mainly significant in the right tail of the return distribution, with negative coefficients, suggesting that higher interest rates are associated with a reduction in the market's upside potential. Finally, changes in the default yield affect the left and right tail of the return distribution in a largely symmetric manner, with a higher default yield being associated with a bigger spread in the conditional return distribution, i.e., thicker left and right tails.

In contrast to the strong evidence of predictability in the tails of the return distribution, there is little evidence of predictability in the center of the return distribution, with only two of the 16 quantile estimates being significant at the 5% level for $\alpha = 0.4, 0.5, 0.6, 0.7$.

As a clearer illustration of the ability of the state variables to predict different parts of the return distribution, Figure 1 plots the quantile regression estimates. The horizontal axis lists quantiles running from 0.05 through 0.95, while coefficient estimates showing the effect of the state variables on the individual quantiles along with standard error bands are listed on the vertical axis. If all parts of the return distribution were unpredictable, all quantile estimates should be insignificant.

---

[3] The linear regression model in Eq. (**??**) where x_t simply shifts the conditional mean of the return distribution emerges from Eq. (**??**) when $\beta_{1,\alpha}$ does not vary across quantiles, i.e. $\beta_{1,\alpha} = \beta_1$ for all α, while $\beta_{2,\alpha} = \beta_{3,\alpha} = 0$:

$$q_\alpha(r_{t+1}|\mathcal{F}_t) = \beta_{0,\alpha} + \beta_1' x_t. \tag{14}$$

In fact, coefficient estimates from the quantile model follow systematic patterns. For example, the coefficient estimates of the default yield are large and negative in the left tail (small quantiles) and large and positive in the right tail (large quantiles). More broadly, most of the quantile estimates are significant in the tails and "shoulders" of the return distribution. Despite having little ability to predict the center (mean) of the return distribution, the economic state variables are thus capable of predicting outcomes in the tails.

As a summary measure of whether a given predictor variable is significant if all quantiles are considered jointly, the last column of Table 2 reports Bonferroni $p-$values. These are robust to arbitrary dependencies across individual quantile estimates. By this criterion, all predictor variables except the T-bill rate are (jointly) significant at the 1% critical level. Three out of four predictor variables are thus helpful in predicting some part of the probability distribution of stock returns.

Finally, using the general quantile model in Eq. (**??**), the last rows in Table 2 show that dynamic effects are important for modeling the quantiles with exception of the center of the return distribution. The lagged quantile is significant for all quantiles except for the median and the lagged absolute value of returns is also significant for all but the center quantiles.

To gain intuition for how these findings translate into variations in the predictive return distribution, Figure 2 shows the quantiles of returns computed under three sets of values for the predictor variables: A middle scenario that sets all variables in Eq. (**??**) at their sample means and two scenarios where each covariate is separately set at its mean plus or minus two standard deviations, while keeping the remaining covariates at their sample means.

For the dividend yield and term spread we clearly see the much bigger impact from a change in the state variable on the left tail, i.e., on downside risk, than on the right tail, i.e., upside potential. Moreover, changes in the dividend yield lead to a parallel shift in the return distribution, consistent with the earlier finding of a mean effect for this variable. Conversely, movements in the term spread and the T-bill rate affect the conditional spread of the return distribution. Increases in the default spread shift the lower quantiles downwards and the upper quantiles upwards, reflecting a greater chance of either large negative or large positive returns.

To see how the distribution forecasts−as measured by the conditional quantiles−evolve over time, Figure 3 plots the 5%, 10%, 50%, 90% and 95% quantiles over the period 1956-2008 based on the dynamic quantile (DQX) and TVM-EGARCHX models. Horizontal lines show the corresponding quantiles for the model that assumes constant quantiles. There is considerable variation over time in the conditional quantiles. Moreover, this variation is highly persistent and much stronger in the tails than at the median. Some patterns in return predictability are clearly volatility driven.

This includes the period following the oil price shocks of 1974/75 and during the shift in the monetary policy regime from 1979 to 1981. Both episodes were associated with highly uncertain market conditions. At other times, e.g., from November 1979 to May 1980, the lower quantiles decline significantly more than the upper quantiles rise, indicating substantial downside risk.

In summary, the empirical results so far indicate that there is significant in-sample return predictability, but the evidence is far weaker for the center of the return distribution (captured by its mean or median value) than in the tails of the return distribution. Clearly, the difficulty many empirical studies have had in establishing predictability of mean returns does not imply that the return distribution cannot be predicted in ways that could potentially be of economic value to investors.

3 Statistical Measures of Return Predictability

To measure the accuracy of the forecasts from the return prediction models, we first report the dominant statistical measure of forecasting performance used in the finance literature, namely the root mean squared forecast error. Next, we consider more general measures of forecasting performance that consider the entire return distribution. These allow us to address whether tails of the return distribution can be predicted and, if so, whether the models perform equally well in the left and right tails. As we shall see, the performance of the return prediction models in the tails provides important clues for whether they add information that is valuable to investors' portfolio selection.

Throughout the analysis we consider the eight forecasting models described in Section 2:

1. Prevailing mean and variance (PMV, Eq. (**??**)).
2. Time-varying mean (TVM, Eq. (**??**)).
3. Constant mean and time-varying volatility (EGARCH, Eq. (**??**)),
4. Time-varying mean and volatility (TVM-EGARCH, Eq. (**??**)).
5. Time-varying mean and volatility with economic predictors (TVM-EGARCHX, Eq. (**??**)).
6. Prevailing quantiles (PQ, Eq. (**??**)).
7. Dynamic quantiles (DQ, Eq. (**??**)).
8. Dynamic quantiles with economic predictors, (DQX, Eq. (**??**)).

The importance of parameter estimation error is gauged by considering both in-sample and out-of-sample forecasts. The former use forecasts from models whose parameters are based on full-sample estimates. In-sample forecasts from the constant distribution models in equations (**??**) and (**??**) are therefore the same each period whereas forecasts from all other approaches vary over time due to changes in the mean, variance or individual quantiles.

Out-of-sample forecasts are based on parameter estimates that use a 30-year rolling window of data to guard against changes in the data generating process. A relatively long estimation window is chosen to reduce the effect of parameter estimation error which would otherwise lead to imprecise estimates of tail probabilities. Out-of-sample forecasts are computed as follows. First, the parameters of the return prediction models are estimated using data from the first 30-year sample, 1926:01-1955:12. One-step-ahead forecasts are then generated for the distribution of returns in 1956:01. The following month our estimates are updated by rolling the estimation window one observation ahead and adding data from 1956:01. The updated model is used to produce return distribution forecasts for 1956:02. This recursive forecasting procedure is repeated up to the end of the sample, generating 636 out-of-sample forecasts for the period 1956:01-2008:12. For comparability purposes the in-sample forecasting results are based on the same sample.

3.1 Root Mean Squared Errors

Table 3 reports root mean squared error (RMSE) values for the eight forecasting models. As expected, the models that allow for a time-varying mean perform best in-sample, irrespective of whether the volatility is kept constant or is allowed to change over time. The worst models by this criterion are the quantile models for which the point forecasts use the conditional median. This is perhaps not surprising since the objective of these models is not to minimize the RMSE or to predict the mean.

A very different picture emerges for the out-of-sample experiment in which the best model is the prevailing mean and variance (PMV) specification followed by the PM-EGARCH and the TVM-EGARCH models. The prevailing quantile model also performs quite well, showing the importance of keeping a simple model when it comes to out-of-sample forecasting based on the RMSE criterion. This criterion focuses on predictability of the mean, however, and so may not adequately reflect the value of return predictions from the perspective of economic investors who are concerned with the fit of the entire return distribution. To address this issue, we next evaluate the return probability forecasts.

3.2 Probability Forecasts

To find out how well our return prediction models capture different parts of the return distribution, we next evaluate the density forecasts. These are computed as follows. The EGARCH models assume that returns are (conditionally) normally distributed with time-varying volatility, and, possibly, a time-varying mean. For the quantile models the distribution of excess returns, $f(r_{t+1}|\mathcal{F}_t)$, is approximated by assuming that this is piecewise uniformly distributed between the quantile forecasts with exponentially decaying tails:

$$f(r_{t+1}|\mathcal{F}_t) = \begin{cases} \frac{1}{\sqrt{2\pi}\tilde{\sigma}_{t+1}} \exp(\frac{-(r_{t+1}-\tilde{\mu}_{t+1})^2}{2\tilde{\sigma}^2_{t+1}}), & \text{if } r_{t+1} \leq q_{0.05,t+1} \\ \frac{0.05}{q_{0.10,t+1}-q_{0.05,t+1}}, & q_{0.05,t+1} \leq r_{t+1} \leq q_{0.10,t+1} \\ \frac{0.1}{q_{\alpha+0.10,t+1}-q_{\alpha,t+1}}, & q_{\alpha,t+1} \leq r_{t+1} \leq q_{\alpha+0.10,t+1} \\ & (\alpha \in [0.10, 0.80]) \\ \frac{0.05}{q_{0.95,t+1}-q_{0.90,t+1}}, & q_{0.90,t+1} \leq r_{t+1} \leq q_{0.95,t+1} \\ \frac{1}{\sqrt{2\pi}\tilde{\sigma}_{t+1}} \exp(\frac{-(r_{t+1}-\tilde{\mu}_{t+1})^2}{2\tilde{\sigma}^2_{t+1}}), & \text{if } r_{t+1} > q_{0.95,t+1} \end{cases} \tag{15}$$

where $q_{\alpha,t+1}$ is the predicted $\alpha-$quantile given information at time t and $\tilde{\mu}_{t+1}$ and $\tilde{\sigma}_{t+1}$ are estimates of the center and dispersion of the return distribution which ensure that the return distribution is continuous at the 5% and 95% quantiles.[4]

Each forecasting model gives rise to a density $f(r_{t+1}|\mathcal{F}_t)$. Provided that the density is correctly specified, the so-called probability integral transforms evaluated at the actual values of excess returns, $u_{t+1} = \int_{-\infty}^{r_{t+1}} f(y|\mathcal{F}_t)dy$ should be independently, identically, and uniformly distributed. Figure 4 provides a visual illustration of this property by plotting histograms for the probability integral transforms (u_{t+1}) computed out-of-sample. For correctly specified models, the frequency of each bin should be approximately one. Bins whose frequency fall above one indicate parts of the return distribution whose likelihood is under-predicted, while bins with frequency below one indicate parts of the distribution whose likelihood is over-predicted. The prevailing mean and variance model (PMV) in particular, but also the time-varying mean and constant variance (TVM) model, overestimate probabilities in both the left and right tails, as indicated by the fact that fewer than expected returns fall in the top and bottom tails. Conversely, these models underestimate

---

[4] Instead of assuming uniform distributions between the individual quantiles, we also considered Gaussian kernels for the probability distribution between the individual quantiles. This approach yields very similar results and so is not reported here.

the likelihood of returns in the center. The TVM-EGARCH and TVM-EGARCHX models fit the return distribution much better overall, with a slight tendency to over-predict returns in the left tail. Finally, the quantile models produce reasonably good distribution forecasts, although the multivariate dynamic quantile (DQX) model seems to be misspecified both in the center and in the right tail.[5]

As a more formal test of how well time-variations in different parts of the return distribution is captured by a particular model, we consider the approach proposed by Christoffersen (1998) applied to the individual quantiles. Define the indicator variable $I_t = \mathbf{1}\{\hat{q}_{\alpha_1,t+1} \leq r_{t+1} < \hat{q}_{\alpha_2,t+1}\}$ which equals one when the actual return falls between the predicted α_1 and α_2 quantiles ($\alpha_2 > \alpha_1$) and otherwise equals zero. Christoffersen shows that if a probability distribution model is correctly specified at these $\alpha-$quantiles, then the sequence I_t, $t = 1, .., T$ should be identically and independently distributed and follow a Bernoulli distribution with parameter $\alpha_2 - \alpha_1$. Furthermore, there should not be any persistence in the dynamics of the indicator variable, so I_t should not be serially correlated. Both properties can be tested through likelihood ratio tests with rejection suggesting that the probability distribution model is misspecified at a particular quantile.

Table 4 shows the outcome of these tests with the unconditional coverage tests in Panel A and the persistence tests in Panel B. The two models that assume a constant volatility, i.e., the prevailing mean and variance (PMV) and time-varying mean (TVM) specifications, are clearly misspecified and get rejected both in the left tail (0.10-0.20), the center (0.50-0.60) and in the right tail (0.90-1). Most notably, these models overpredict the thickness of the right tail and underpredict the center of the return distribution. The multivariate dynamic quantile model (DQX) is also rejected in the left shoulder (0.30-0.50) and in the right tail (0.90-1). Conversely, there is little evidence of misspecification for the two EGARCH models that allow for a time-varying mean and volatility.

Panel B of Table 4 reports a test based on the probability of return observations repeatedly falling at a particular point of the return distribution, $prob(\hat{q}_{\alpha_1,t+1} \leq r_{t+1} < \hat{q}_{\alpha_2,t+1} | \hat{q}_{\alpha_1,t} \leq r_t < \hat{q}_{\alpha_2,t})$. The constant volatility models have far too high a probability of drawing repeated outliers from the left tail (0-0.10) and too low a probability of repeated observations in the right tail (0.90-1) and so get rejected in the tails by this test. While these probabilities are also too high in the left tail of the EGARCH models, they are substantially smaller than those found for the PMV model

---

[5] The tails, as considered here, cover the top or bottom 10-20% of the return distribution. While the return prediction models to varying degrees tend to overpredict these tails, they all underpredict the extreme left and right tails defined as the top or bottom one percent of the return distribution. Returns this far out in the tails reflect events such as October 1987 which, unsurprisingly, none of the models can accommodate and, since they affect all models, do not affect the models' relative performance.

in particular.

3.3 Pairwise Comparison of Distribution Forecasts

The probability tests reported in Table 4 measure whether the individual models are correctly specified. It is of separate interest to ask whether some models are better than others. To analyze the *relative* performance of the density forecasts, we use the weighted likelihood ratio test statistic of Amisano and Giacomini (2007):

$$t = \frac{\overline{WLR}_T}{\widehat{\sigma}_T/\sqrt{T}}, \tag{16}$$

where $\overline{WLR}_T = T^{-1}\sum_{t=0}^{T-1} WLR_{t+1}$ is the average weighted likelihood ratio for an out-of-sample period of T observations and $\widehat{\sigma}_T$ is an estimator of its variance. The weighted likelihood ratio (WLR_{t+1}) is the weighted average difference between log scores of two competing density forecasts evaluated at the actual return:

$$WLR_{t+1} = w(\bar{r}_{t+1})(\log f_{t+1}(r_{t+1}) - \log g_{t+1}(r_{t+1})), \tag{17}$$

where $w(\bar{r}_{t+1})$ is a weight function evaluated at the standardized return at time $t+1$, $\bar{r}_{t+1} = (r_{t+1} - \hat{\mu}_{t+1})/\hat{\sigma}_{t+1}$, and $\hat{\mu}_{t+1}$ and $\hat{\sigma}_{t+1}$ are the estimated mean and standard deviation of r_{t+1}. Using different weighting schemes, different parts of the return distribution can be analyzed. We consider four weighting schemes:

1. Full distribution: $w(x) = 1$;

2. Center of the distribution: $w(x) = \phi(x)$, ϕ standard normal density function;

3. Left tail of the distribution: $w(x) = 1 - \Phi(x)$; $\Phi(x)$ normal distribution function.

4. Right tail of the distribution: $w(x) = \Phi(x)$.

Table 5 reports pairwise out-of-sample comparisons of model fit using these weighting schemes. Specifically, the table presents the predictive performance of the row models measured against the performance of the column models. Positive numbers suggest that the row model is best, while negative numbers show that the column model dominates. The two constant volatility models (PMV and TVM) perform very poorly and get rejected against all the EGARCH specifications in the center of the return distribution as well as in the right tail of the return distribution for the TVM model. Interestingly, the dynamic quantile models perform relatively poorly out-of-sample, particularly the more heavily parameterized multivariate model (DQX). This shows the importance of parameter estimation error. The best performing models are the EGARCH specifications, particularly the simple model that allows for a time-varying mean but excludes the predictor variables from the volatility equation (TVM-EGARCH).

We conclude the following from the empirical analysis up to this point. First, consistent with the existing literature (e.g., Welch and Goyal (2008)) we find that the simple prevailing mean and variance model produces the lowest out-of-sample RMSE-values and thus generates the best forecasts of the mean equity premium. However, this model also has severe shortcomings and, most notably, significantly overestimates the tails of the return distribution and underestimates the center. Models that allow for a time-varying mean and variance of returns (most notably EGARCH specifications) produce noisier forecasts of mean returns, but generate better forecasts of the probability distribution of stock returns.

4 Economic Measures of Return Predictability

The empirical evidence so far suggests that there is significant statistical evidence of time-variation in the return distribution, particularly in the tails where the constant return distribution model gets strongly rejected. This section asks if such predictability is sufficiently important to be of economic value to investors by considering the usefulness of predictability from the perspective of a small, risk averse investor with no market impact who chooses portfolio weights based on the probability distribution forecasts.

4.1 Portfolio Selection

Consider an investor who at time t allocates $w_t W_t$ of total wealth to stocks and the remainder, $(1 - w_t)W_t$ to a risk-free asset, where $W_t = 1$ is the initial wealth. The investor's wealth at time $t + 1$ is

$$W_{t+1} = (1 - \omega_t)\exp(r^f_{t+1}) + \omega_t \exp(r^f_{t+1} + r_{t+1}),$$

where r_{t+1} is the stock return in excess of the risk-free rate, r^f_{t+1}. The investor is assumed to have constant relative risk aversion (CRRA) preferences over next period's wealth,

$$U(W_{t+1}) = \frac{W_{t+1}^{1-\gamma}}{1-\gamma}, \tag{18}$$

where γ is the coefficient of relative risk aversion. Portfolio weights for period t can be obtained as the solution to the following optimization problem:

$$w_t^* = \arg\max_{w_t} E_t[U(W_{t+1})], \tag{19}$$

where $E_t[\cdot]$ denotes the conditional expectation based on the investor's information at time t. The investment horizon is set to one period and any intertemporal hedging component in the investor's portfolio choice is ignored. Hence we assume that the investor solves Eq. (**??**), holds the optimal portfolio for one period and then reoptimizes the portfolio weights the following period based on any new information. The portfolio optimization problem in Eq. (**??**) can thus be written as

$$w_t^* = \arg\max_{w_t} \int \frac{1}{1-\gamma}((1-\omega_t)\exp(r_{t+1}^f) + \omega_t \exp(r_{t+1}^f + r_{t+1}))^{1-\gamma} f(r_{t+1}|\mathcal{F}_t) dr_{t+1}, \tag{20}$$

where $f(r_{t+1}|\mathcal{F}_t)$ is the conditional probability distribution of future excess returns. Following Kandel and Stambaugh (1996) and Geweke (2001), we restrict $w \in [0, 0.99]$ to ensure that the expected utility is finite.

In special cases such as log-normally distributed returns and CRRA utility, the portfolio allocation can be derived in closed form. Suppose that the conditional distribution of log excess returns is normal with predicted mean and variance $\widehat{\mu}_{t+1}$ and $\widehat{\sigma}_{t+1}^2$. Campbell and Viceira (2001) derive the following log-linearized approximation to the investor's wealth at time $t+1$:

$$\ln(W_{t+1}) \approx \ln(1 + r_{t+1}^f) + \omega_t(\ln(1 + r_{t+1}^s) - \ln(1 + r_{t+1}^f)) + \frac{1}{2}\omega_t(1-\omega_t)\widehat{\sigma}_{t+1}^2, \tag{21}$$

from which the approximate optimal portfolio weight for an investor with CRRA utility is obtained:

$$\omega_t^* \approx \frac{\widehat{\mu}_{t+1} + \widehat{\sigma}_{t+1}^2/2}{\gamma \widehat{\sigma}_{t+1}^2}. \tag{22}$$

This equation can be used for the optimal portfolio allocation of a CRRA investor under the assumption of the PMV, TVM or EGARCH models for returns in equations (**??**) - (**??**).

To solve Eq. (**??**) under the quantile models we use the conditional return distribution in Eq. (**??**) and compute the resulting integrals numerically.

The performance of optimal portfolios based on the various density forecasts is compared through the associated certainty equivalent return (CER) estimate under CRRA utility:

$$CER = \left((1-\gamma)T^{-1}\sum_{t=1}^{T} U(W_t^*)\right)^{1/(1-\gamma)} - 1, \tag{23}$$

where T is the total number of observations in the out-of-sample period, $1/T\sum_{t=1}^{T} U(W_t^*)$ is the mean realized utility and the realized utility for period t can be calculated through the sequence of optimal portfolio weights, ω_{t-1}^*.

4.2 Empirical Results

Table 6 presents certainty equivalent return estimates in addition to the Sharpe ratio, mean, standard deviation, skew and kurtosis of portfolio returns for the in-sample and out-of-sample forecast experiments. We assume CRRA preferences with a coefficient of relative risk aversion, $\gamma = 5$. In the in-sample experiment (Panel A), the multivariate dynamic quantile (DQX) model produces the highest certainty equivalent return of 8.7% per annum, followed by the time-varying mean (TVM) model (8.2%) and the TVM-EGARCHX and TVM-EGARCH models (8.1% and 7.9%, respectively). The prevailing mean and variance model (PMV), PQ and PM-EGARCH models generate somewhat lower certainty equivalent returns of 6.8%. Allowing for a time-varying mean in particular thus seems to lead to better in-sample economic performance.

Turning to the out-of-sample results reported in Panel B, the time-varying mean EGARCH models now dominate and generate certainty equivalent returns in excess of 7% per annum. The models that allow for either a time-varying mean (TVM, 6.3%) or a time-varying volatility (PM-EGARCH, 5.7%) perform better than the prevailing mean model (PM, 5.5%). All quantile models perform quite poorly out-of-sample with certainty equivalent returns in the range 5.0% - 5.6%, suggesting that parameter estimation error is a particular concern for these models.

We conclude the following from these results. First, predictable time-variation in the distribution of stock returns seems to be sufficiently strong even out-of-sample to be of economic value to a risk averse investor who uses one of the TVM-EGARCH models which generate out-of-sample certainty equivalent returns that are 150 basis points per year higher than those associated with the constant return distribution models. Second, comparing the in-sample and out-of-sample results, parameter estimation error systematically reduces the certainty equivalent returns ranging from 70 basis points for the TVM-EGARCH model to 210 basis points for the dynamic quantile (DQX) model whose parameters are most difficult to estimate precisely and so are surrounded by the greatest sampling errors.

4.3 Pairwise Comparison of Risk-adjusted Performance

As in the case of the probability distribution forecasts, we can again ask if the risk-adjusted returns associated with some forecasting models are significantly higher than those associated with others, i.e., whether the relative return performance is significantly better for some models. To this end we use the non-parametric stochastic discount factor approach of Almeida and Garcia (2008). This approach accounts for higher moments of the stochastic discount factor which is important here in view of the evidence of strongly non-normally distributed stock returns. In a first stage, the stochastic discount factor is backed out from data on a set of underlying risk factors. In a second stage, returns on a test portfolio get compared to a set of competing (benchmark) returns using the sample analog of an Euler equation. If the sample estimate of the product of the differential portfolio returns and the stochastic discount factor is positive and significant, we conclude that the test portfolio provides higher risk-adjusted returns than its competitor.

Specifically, we first estimate the stochastic discount factor from the conventional asset pricing model, $E(m_t R_t - 1) = 0$, where 1 and 0 are vectors of ones and zeros, respectively, and R_t is a vector of gross returns on a set of underlying risk factors that are priced by the stochastic discount factor, m_t. The sample analog of this equation is

$$\frac{1}{T}\sum_{t=1}^{T} m_t(R_t - \frac{1}{\overline{m}}) = 0, \tag{24}$$

where $\overline{m} = E(m_t)$ is the unconditional mean of the stochastic discount factor.

The sample estimate of the stochastic discount factor implied by the Almeida-Garcia (2008) approach is

$$\widehat{m}_t = T \times \overline{m} \times \frac{(1 - \psi\widehat{\varphi}'(R_t - \frac{1}{\overline{m}}))^{\frac{1}{\psi}}}{\sum_{t=1}^{T}(1 - \psi\widehat{\varphi}'(R_t - \frac{1}{\overline{m}}))^{\frac{1}{\psi}}}. \tag{25}$$

Here ψ can be interpreted as the parameter of an investor with HARA preferences of the form $U(W) = -\frac{1}{1+\psi}(1 - \psi W)^{\frac{1+\psi}{\psi}}$. Moreover, $\widehat{\varphi}$ can be obtained from the solution to an empirical likelihood problem.[6]

The performance of a pair of portfolios associated with competing return prediction models can now easily be compared using the sample analog of the Euler equation for the estimated stochastic discount factor, $\widehat{m}_t$, and the differential returns on the portfolios, $R_{1t} - R_{2t}$:

$$\widehat{\alpha} = \frac{1}{T}\sum_{t=1}^{T} \widehat{m}_t(R_{1t} - R_{2t}), \tag{26}$$

where $\widehat{\alpha}$ measures the risk-adjusted return of portfolio 1 in excess of the benchmark return (portfolio 2).

4.3.1 Empirical Results

Monthly returns on the S&P500 were used to estimate the implied stochastic discount factor. The hyperparameter ψ was set to one-half, although the results were found not to be sensitive to this choice. The unconditional mean of the stochastic discount factor, $\overline{m}$, was set to the reciprocal of the mean gross return on the risk free asset during the sample period 1956:01-2008:12. Model

---

[6] Almeida and Garcia establish conditions under which

$$\widehat{\varphi} = \sup_{\varphi \in \Upsilon} \frac{\overline{m}^{1+\psi}}{T} \sum_{t=1}^{T} -\frac{1}{1+\psi}\left(1 - \psi\varphi'\left(R_t - \frac{1}{\overline{m}}\right)\right)^{\frac{1+\psi}{\psi}},$$

where φ is constrained to the set $\Upsilon = \{\varphi \in \mathbb{R}^k, \text{s.t. } t = 1, 2, \ldots, T : (1 - \psi\varphi'(R_t - \frac{1}{\overline{m}})) > 0\}$ and ψ is a hyperparameter of the so-called Cressie Read discrepancy function $\frac{x^{1+\psi}-1}{(1+\psi)\psi}$.

comparisons were performed pair-wise and p-values for the risk-adjusted returns associated with the sample Euler equation in Eq. (**??**) calculated using the stationary bootstrap (Politis and Romano (1994)) on the alpha estimates.[7]

First consider the in-sample evidence reported in Panel A of Table 7. Once again, positive numbers indicate that the row model is best, while negative numbers suggest that the column model is best. The prevailing mean and variance (PMV) model is worse than all but the prevailing quantile model, with a risk-adjusted return significantly lower than that associated with the TVM model, the two EGARCH models with time-varying means (TVM-EGARCH and TVM-EGARCHX) and the multivariate dynamic quantile model (DQX). These models dominate the PMV model by between 170 and 250 basis points per annum in risk-adjusted performance. Overall, the best model as measured by the relative risk-adjusted return performance is the DQX model which dominates all other specifications by at least 40 basis points per annum, significantly outperforming four of the seven other models at the 1% critical level.

Turning to the out-of-sample results shown in Panel B, the prevailing mean and variance model (PMV) lose to the EGARCH models. In fact, the risk-adjusted returns of the TVM-EGARCH and TVM-EGARCHX models are 200 basis points per year higher than that of the prevailing mean and variance model. The time-varying mean model (TVM) is also dominated by the TVM-EGARCH specification. Overall, the best model is the TVM-EGARCHX specification which outperforms all other models in pair-wise comparisons, five of them at the 10% critical level, with the TVM-EGARCH model a close second.

4.4 Relating Statistical and Economic Measures of Forecasting Performance

Our empirical results suggest that return prediction models with constant mean perform best when it comes to forecasting performance as measured by the out-of-sample mean squared forecast error. However, these models perform poorly when measured by the economic value of their forecasts used to guide portfolio selection. The opposite finding holds for models that allow for time-varying mean and volatility. These models also perform better than constant distribution models when judged by their ability to predict the probability distribution of returns measured by the density prediction test. This raises an important question, namely how closely related different statistical and economic measures of forecasting performance are.

The ranking of the return prediction models according to the certainty equivalent returns or

---

[7] Specifically, we generate 999 random samples of 636 observations from the estimated relative performance measures using the centered stationary bootstrap approach and a block size determined by a geometric distribution with a parameter of 0.05.

mean returns in Table 6 turns out to be much closer to the ranking of the models based on the full density forecast comparison test in Table 5 than they are to the mean squared errors. To show this, Figure 5 provides scatter plots of either the (out-of-sample) RMSE-values of the eight models (left plots) or the value of the Amisano-Giacomini (2007) density comparison test for the full return distribution (right plots) against the associated certainty equivalent returns (top windows), the Sharpe ratios (middle windows) or the mean returns (bottom windows). If the RMSE measure of forecasting performance were informative of the economic value of the associated forecasts, a strong negative correlation between this measure and the economic performance measure should emerge, i.e., higher RMSE-values (poorer fit) should be associated with worse return performance. Conversely, higher values of the density forecast comparison test indicate better fitting models, and so we would expect a positive relation between this test and the economic performance measures.

Using the Spearman rank correlation test, we find only mild evidence of a negative rank correlation between the RMSE values and either certainty equivalent returns, Sharpe ratios, or mean returns, with correlations ranging between -0.47 and -0.04. The associated p-values range from 0.23 to 0.91. In contrast, there is a much stronger rank correlation (ranging from 0.60 to 0.83) between the predictive density LR test and either certainty equivalent returns, Sharpe ratios, or mean returns. Moreover, these rank correlations are statistically significant at either the 10% or 5% level with p-values ranging from 0.01 to 0.07.

The close relation between the accuracy of the forecasts of the return probability distribution and the economic value of such forecasts is explained by considering how stock holdings map into the cost of using a misspecified model for the return probability distribution. The prevailing mean model severely overestimates the right tail of the return distribution with a resulting high average allocation to stocks (70% in our sample). This is much higher than the 50% average allocation to stocks under the TVM-EGARCH models that better capture the full return distribution. The larger stock holdings under the constant mean and variance model gives rise to low utility during periods with large negative returns.

Our results show that very different conclusions on return predictability emerge depending on which loss function is used to measure predictive accuracy. This finding is part of a more general point.[8] As argued by Granger and Machina (2006), when predictions are used for decision making such as portfolio selection, the preferences and constraints of the decision maker enter into the evaluation and comparisons of the forecasts and naturally leads to decision-based loss-functions. Granger and Machina show that the loss function underlying mean squared error loss implies

---

[8] Christoffersen and Jacobs (2004) also emphasize the importance of the loss function in the context of option valuation.

highly restrictive and unrealistic properties of the underlying preferences. Our results suggest that predictive accuracy measured under squared error loss is not a good indicator of the economic value of return predictability for investors with power utility.

5 Conclusion

This paper shows that measures such as root mean squared forecast errors contain little information when it comes to assessing whether return prediction models that allow for a time-varying return distribution help or hurt investors. In fact, there is only a weak correlation between how well a model performs in terms of its out-of-sample root mean squared forecast error and the economic value of using its forecasts in portfolio allocation. More comprehensive measures of predictive accuracy that consider the full return distribution are needed to understand why some forecasting models succeed in adding economic value to investors, while others fail.

Empirically, we find evidence that allowing for predictability in both the mean and variance of returns leads to more accurate forecasts of the probability distribution of stock returns although it does not lead to better forecasts of mean returns. Our analysis shows that prediction models that assume constant mean returns tend to overpredict the right tail of the return distribution which leads to too large investments in stocks and can be costly to risk averse investors. Return prediction models with a time-varying mean and variance lead to better estimates of the tails of the return distribution and suffer less from surprises in the form of repeated unanticipated outliers in returns compared with a constant return distribution model. Using the time-varying distribution forecasts in asset allocation results in a risk-adjusted return performance close to 2% per annum higher than if a constant mean and volatility model is used to generate out-of-sample forecasts.

Our analysis suggests that the debate on return predictability has focused too narrowly on statistical measures of forecast precision such as root mean squared forecast errors or R^2 and that important insights can be learned by focusing instead on more comprehensive measures of predictive accuracy of the full return distribution that are more closely related to the economic value of allowing for a time-varying probability distribution. Despite their failure to produce more accurate forecasts of mean returns than the simple constant-mean benchmark of Welch and Goyal (2008), return forecasting models that allow for time-varying mean and volatility do seem to add value over simple benchmarks that ignore such return predictability.

References

[1] Almeida, C. and R. Garcia, 2008, Empirical Likelihood Estimators for Stochastic Discount Factors. Working paper, Getulio Vargas Foundation and Edhec, Nice.

[2] Amisano, G., and R. Giacomini, 2007, Comparing Density Forecasts via Weighted Likelihood Ratio Tests. Journal of Business and Economic Statistics 25, 177-190.

[3] Ang, A. and G. Bekaert, 2007, Stock Return Predictability: Is it There? Review of Financial Studies 20, 651-707.

[4] Bali, T. and N. Cakici, 2010, World Market Risk, Country-Specific Risk and Expected Returns in International Stock Markets. Journal of Banking and Finance 34(6), 1152-1165.

[5] Bali, T., H. Mo, and Y. Tang, 2008, The Role of Autoregressive Conditional Skewness and Kurtosis in the Estimation of Conditional VaR. Journal of Banking and Finance 32(2), 269-282.

[6] Bossaerts, P. and P. Hillion, 1999, Implementing Statistical Criteria to Select Return Forecasting Models: What do we Learn? Review of Financial Studies 12(2), 405-428.

[7] Campbell, J.Y., 1987, Stock Returns and the Term Structure. Journal of Financial Economics 18(2), 373-399.

[8] Campbell, J.Y., and R., Shiller, 1988, The Dividend-Price Ratio, Expectations of Future Dividends and Discount Factors, Review of Financial Studies 1, 195-227.

[9] Campbell, J.Y. and S.B. Thompson, 2008, Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? Review of Financial Studies 21, 1509-1532.

[10] Campbell, J.Y. and L.M. Viceira, 2001, Who Should Buy Long-Term Bonds? American Economic Review 91, 99-127.

[11] Christoffersen, P.F., 1998, Evaluating Interval Forecasts. International Economic Review 39, 841-862.

[12] Christoffersen, P.F. and K. Jacobs, 2004, The Importance of the Loss Function in Option Valuation. Journal of Financial Economics 72, 291-318.

[13] Cochrane, J.H., 2008, The Dog that did not Bark: A Defense of Return Predictability. Review of Financial Studies 21, 1533-1576.

[14] Dangl, T. and M. Halling, 2008, Predictive Regressions with Time-Varying Coefficients. Working Paper, Vienna University of Technology.

[15] Engle, R.F., E. Ghysels and B. Sohn, 2007, On the Economic Sources of Stock Market Volatility. Manuscript, New York University and University of North Carolina.

[16] Engle, R. F. and S. Manganelli, 2004, CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. Journal of Business and Economic Statistics 22(4), 367-381.

[17] Fama, E.F. and K.R. French, 1988, Dividend Yields and Expected Stock Returns. Journal of Financial Economics 22(1), 3-25.

[18] Fama, E.F. and K.R. French, 1989, Business Conditions and Expected Returns on Stocks and Bonds. Journal of Financial Economics 25(1), 23-49.

[19] Ferson, W., 1990, Are the Latent Variables in Time-Varying Expected Returns Compensation for Consumption Risk?, Journal of Finance, 45, 397-429.

[20] Ferson, W., and C., Harvey, 1993, The Risk and Predictability of International Equity Returns, Review of Financial Studies, 6, 527-566.

[21] Foresi, S. and F. Peracchi, 1995, The Conditional Distribution of Excess Returns: An Empirical Analysis. Journal of the American Statistical Association 90, 451-466.

[22] Geweke, J., 2001, A Note on Some Limitations of CRRA Utility. Economics Letters 71, 341-345.

[23] Ghysels, E., P. Santa-Clara and R. Valkanov, 2006, Predicting Volatility: How to get the Most out of Return Data Sampled at Different Frequencies. Journal of Econometrics 131, 59-95.

[24] Glosten, L.R., R. Jagannathan, and D.E. Runkle, 1993, On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. Journal of Finance 48, 1779-2801.

[25] Granger, C.W.J., and M.J. Machina, 2006, Forecasting and Decision Theory. Pages 81-98 in G. Elliott, C.W.J. Granger and A. Timmermann (eds.), Handbook of Economic Forecasting, vol. 1. North Holland.

[26] Johannes, M., A. Korteweg and N. Polson, 2009, Sequential Learning, Predictive Regressions, and Optimal Portfolio Returns. Mimeo, Columbia University.

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

[28] Koenker, R. and G. Bassett, 1978, Regression Quantiles. Econometrica 46, 33-50.

[29] Komunjer, I., 2005, Quasi-Maximum Likelihood Estimation for Conditional Quantiles. Journal of Econometrics 128, 137-164.

[30] Lettau, M. and S. Ludvigsson, 2001, Consumption, aggregate wealth, and expected stock returns. Journal of Finance 56, 815-850.

[31] Lettau, M. and S. van Nieuwerburgh, 2008, Reconciling the Return Predictability Evidence. Review of Financial Studies 21, 1607-1652.

[32] Merton, R., 1980, On Estimating the Expected Return on the Market: An Exploratory Investigation. Journal of Financial Economics 8, 323-361.

[33] Paye, B., 2010, Which Variables Forecast Aggregate Stock Market Volatility? Working Paper, Rice University.

[34] Paye, B. and A. Timmermann, 2006, Instability of Return Prediction Models. Journal of Empirical Finance 13, 274-315.

[35] Pesaran, M.H. and A. Timmermann, 1995, Predictability of Stock Returns: Robustness and Economic Significance. Journal of Finance 50(4), 1201-1228.

[36] 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.

[37] Politis, D.N., and J.P. Romano, 1994, The Stationary Bootstrap. Journal of the American Statistical Association 89, 1303-1313.

[38] Schwert, G.W., 1989, Why does Stock Market Volatility Change over Time? Journal of Finance 44, 1115-1153.

[39] Welch, I. and A. Goyal, 2008, A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. Review of Financial Studies 21, 1455-1508.

Table 1: Coefficient Estimates from Linear Prediction and EGARCH Models

	TVM	PM-EGARCH	TVM-EGARCH	TVM-EGARCHX
		Mean Equation		
Dividend Yield	0.9827**	-	0.7157**	0.8202**
T-bill Rate	-0.0454	-	-0.1451***	-0.0661
Term Spread	0.1266	-	-0.0301	0.1327
Default Yield	0.0706	-	0.4905	-0.3267
		Variance Equation		
Dividend Yield	-	-	-	-24.1306***
T-bill Rate	-	-	-	-3.9153***
Term Spread	-	-	-	-13.4507***
Default Yield	-	-	-	79.0913***
$\|\varepsilon_{t-1}/\sigma_{t-1}\|$	-	0.2496***	0.2494***	0.0420
$\varepsilon_{t-1}/\sigma_{t-1}$	-	-0.0670***	-0.0719***	-0.2983***
$log(\sigma^2_{t-1})$	-	0.9696***	0.9641***	0.2240*

Note: The coefficient estimates are based on estimation of linear prediction and EGARCH models over the sample 1926:01-2008:12. The OLS estimates are based on a multivariate regression of monthly S&P 500 excess returns on lagged values of the predictor variables listed in the rows. The EGARCH models assume conditionally normally distributed return innovations and are based on maximum likelihood estimation with the lagged predictor variables listed in the rows included as additional variables in the mean and/or variance equations. The coefficient estimates for the dividend yield have been multiplied by 100. All errors are corrected for heteroskedasticity and autocorrelation. * significant at the 10% level. ** significant at the 5% level. *** significant at the 1% level

Table 2: Coefficient Estimates for Quantile Prediction Models

	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	Bonferonni
Dividend Yield	2.3628***	2.5109***	1.1027**	1.1063**	0.6594	0.5658	0.9091**	0.8971	0.8433**	0.2517	-0.3155	0.0000
T-bill Rate	0.2324	0.2420**	0.1308	-0.0083	-0.0929	-0.1250**	-0.1014	-0.1379	-0.1764**	-0.2351**	-0.4492**	0.1210
Term Spread	1.2893***	1.0107***	0.4726***	0.2911*	0.1136	0.0726	0.0320	-0.1088	-0.2005	-0.4429	-0.8187**	0.0000
Default Yield	-5.4976***	-4.6062***	-2.7628***	-1.6823***	-0.6685	-0.2348	0.3365	0.8406	2.0890***	3.6335***	6.8512***	0.0000
Constant	-0.004	-0.002	-0.001	0.000	0.001	0.007***	0.002**	0.003***	0.002***	0.003***	0.002	0.0010
$q_{\alpha,t-1}$	0.840***	0.889***	0.863***	0.925***	0.916***	0.156	0.830***	0.811***	0.874***	0.869***	0.858***	0.0000
$\|r_{t-1}\|$	-0.232***	-0.122***	-0.087***	-0.029***	-0.020	0.032	0.039*	0.084***	0.094***	0.134***	0.241***	0.0000

Note: For each quantile $\alpha = \{0.05, 0.10, \ldots, 0.90, 0.95\}$ the table reports the slope coefficients of lagged predictor variables obtained from quasi-maximum likelihood estimation of the multivariate linear quantile model over the sample 1926:01-2008:12. The dependent variable is the monthly excess return on the S&P 500 index. The significance of the coefficient estimates is based on bootstrapped standard errors using the stationary bootstrap with 1000 replications. The last three rows report the coefficient estimates obtained from quasi-maximum likelihood estimation of the dynamic quantile model without any predictor variables. The final column lists Bonferroni p-values for a joint test across all quantiles that the slope coefficients in the quantile model are equal to zero. The coefficient estimates for the dividend yield have been multiplied by 100. * significant at the 10% level. ** significant at the 5% level. *** significant at the 1% level

Table 3: Root Mean Squared Forecast Errors

	In-sample	Out-of-sample
PMV	4.221	4.247
TVM	4.193	4.323
PM-EGARCH	4.221	4.255
TVM-EGARCH	4.198	4.270
TVM-EGARCHX	4.200	4.286
PQ	4.232	4.278
DQ	4.225	4.302
DQX	4.208	4.397

Note: This table presents in-sample and out-of-sample root mean squared forecast errors over the sample period 1956:01 - 2008:12. All values have been multiplied by 100. The out-of-sample results are obtained by estimating the models recursively using a rolling window with 30 years of observations. The prevailing mean and variance (PMV) model assumes constant mean and variance. The time-varying mean (TVM) model allows the mean to vary over time but assumes constant variance. The PM-EGARCH model assumes a constant mean and a time-varying variance, while the TVM-EGARCH and TVM-EGARCHX models allow for both time-varying mean and variance, with the latter model including the predictor variables in both the mean and variance equations. The prevailing quantile (PQ) model assumes constant quantiles, the dynamic quantile (DQ) model allows for time-varying quantiles, while the DQX model allows the quantiles to depend on their own lagged values as well as on the economic predictor variables. * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Table 4: Specification Tests for Probability Distribution Forecasts

(a) Unconditional Coverage Tests

	0.00-0.10	0.10-0.20	0.20-0.30	0.30-0.40	0.40-0.50	0.50-0.60	0.60-0.70	0.70-0.80	0.80-0.90	0.90-1.00
PMV	0.094	0.072**	0.112	0.115	0.093	0.143***	0.119	0.115	0.077**	0.060***
TVM	0.086	0.077**	0.090	0.094	0.105	0.134***	0.113	0.116	0.104	0.080*
PM-EGARCH	0.115	0.091	0.101	0.101	0.082	0.127**	0.115	0.107	0.093	0.069***
TVM-EGARCH	0.115	0.082	0.086	0.090	0.099	0.110	0.112	0.108	0.099	0.099
TVM-EGARCHX	0.108	0.079*	0.088	0.082	0.093	0.110	0.108	0.121*	0.104	0.107
PQ	0.102	0.101	0.105	0.112	0.088	0.116	0.118	0.094	0.074**	0.090
DQ	0.110	0.101	0.115	0.099	0.090	0.104	0.113	0.093	0.086	0.090
DQX	0.108	0.085	0.110	0.064***	0.127**	0.099	0.090	0.102	0.085	0.129**

(b) Conditional Coverage Tests

	0.00-0.10	0.10-0.20	0.20-0.30	0.30-0.40	0.40-0.50	0.50-0.60	0.60-0.70	0.70-0.80	0.80-0.90	0.90-1.00
PMV	11.908***	8.598**	5.899*	2.033	4.383	13.705***	2.913	2.033	5.506*	14.307***
TVM	9.157**	5.331*	1.287	5.041*	0.570	7.956**	7.618**	2.073	1.092	4.640*
PM-EGARCH	3.529	5.730*	0.650	2.952	2.681	7.867**	1.748	0.572	1.837	10.259***
TVM-EGARCH	6.882**	3.146	1.543	2.182	0.354	1.188	3.523	0.769	0.323	2.649
TVM-EGARCHX	1.701	8.935**	1.233	3.146	2.059	3.598	1.088	4.141	0.999	2.689
PQ	10.590***	4.986*	0.632	1.163	1.233	2.927	2.521	0.528	5.486*	1.150
DQ	6.432**	4.986*	2.033	0.228	1.287	1.177	7.618**	1.101	1.889	2.182
DQX	11.195***	2.344	1.442	10.989***	5.919*	0.780	0.977	0.798	1.960	6.074**

Note: This table presents unconditional coverage probabilities (Panel a) and conditional coverage test statistics (Panel b) evaluated at different intervals of the return distribution. The significance levels in Panel a and Panel b are based on the unconditional and conditional likelihood ratio statistics from Christoffersen (1998), respectively. If a quantile forecast is correctly specified, then the indicator function for realized returns in the interval forecasts should be independently and identically distributed with a success rate equal to 0.10, i.e., the probability of the interval. A significant value suggests that the quantile forecast is misspecified. All forecasts are out-of-sample and cover the period 1956:01-2008:12. Forecasts are obtained by estimating the models recursively using a rolling window with the most recent 30 years of observations. The models are described in the caption to Table 3. * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Table 5: Pairwise Density Forecast Comparisons

(a) Full Distribution

	PMV	TVM	PM-EGARCH	TVM-EGARCH	TVM-EGARCHX	PQ	DQ	DQX
PMV	-							
TVM	-0.0126	-						
PM-EGARCH	0.0379*	0.0505*	-					
TVM-EGARCH	0.0450**	0.0576**	0.0071	-				
TVM-EGARCHX	0.0149	0.0275	-0.0230	-0.0301*	-			
PQ	0.0112	0.0237	-0.0268	-0.0339	-0.0037	-		
DQ	-0.0168	-0.0042	-0.0547*	-0.0618**	-0.0317	-0.0280	-	
DQX	-0.1547***	-0.1421***	-0.1926***	-0.1997***	-0.1696***	-0.1658***	-0.1379***	-

(b) Center of the Distribution

	PMV	TVM	PM-EGARCH	TVM-EGARCH	TVM-EGARCHX	PQ	DQ	DQX
PMV	-							
TVM	-0.0073***	-						
PM-EGARCH	0.0199***	0.0272***	-					
TVM-EGARCH	0.0226***	0.0299***	0.0027	-				
TVM-EGARCHX	0.0200***	0.0272***	0.0000	-0.0027	-			
PQ	0.0145***	0.0218***	-0.0054	-0.0081*	-0.0054	-		
DQ	0.0067	0.0140**	-0.0132**	-0.0159**	-0.0132**	-0.0078	-	
DQX	-0.0305***	-0.0233**	-0.0504***	-0.0532***	-0.0505***	-0.0451***	-0.0373***	-

(c) Left Tail of the Distribution

	PMV	TVM	PM-EGARCH	TVM-EGARCH	TVM-EGARCHX	PQ	DQ	DQX
PMV	-							
TVM	0.0178**	-						
PM-EGARCH	0.0007	-0.0171	-					
TVM-EGARCH	0.0293*	0.0115	0.0286**	-				
TVM-EGARCHX	0.0161	-0.0017	0.0154	-0.0132	-			
PQ	0.0005	-0.0174*	-0.0002	-0.0288*	-0.0156	-		
DQ	-0.0149	-0.0327**	-0.0156	-0.0442**	-0.0310	-0.0154	-	
DQX	-0.0730***	-0.0909***	-0.0737***	-0.1023***	-0.0891***	-0.0735***	-0.0581**	-

(d) Right Tail of the Distribution

	PMV	TVM	PM-EGARCH	TVM-EGARCH	TVM-EGARCHX	PQ	DQ	DQX
PMV	-							
TVM	-0.0304***	-						
PM-EGARCH	0.0372***	0.0676***	-					
TVM-EGARCH	0.0157	0.0462***	-0.0215**	-				
TVM-EGARCHX	-0.0012	0.0292*	-0.0384**	-0.0169**	-			
PQ	0.0107	0.0411***	-0.0265*	-0.0050	0.0119	-		
DQ	-0.0019	0.0285**	-0.0391**	-0.0176	-0.0007	-0.0126	-	
DQX	-0.0816***	-0.0512**	-0.1189***	-0.0974***	-0.0805***	-0.0923***	-0.0797***	-

Note: This table presents the Weighted Likelihood Ratio test statistic of Amisano and Giacomini (2007) for pairwise comparison of the performance of two density forecasts. Models in the rows are compared against models in the columns. A positive number suggests that the model in the row provides a better density forecast than the model in the column, while a negative number suggests the reverse. All results are out-of-sample and cover the period 1956:01-2008:12. Forecasts are obtained by estimating the models recursively using a rolling window with the most recent 30 years of observations. The individual models are described in the caption to Table 3. * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Table 6: Portfolio Return Performance for each of the Density Forecasting Models

(a) In-Sample Results

	CER	Sharpe Ratio	Mean	St. Dev.	Skewness	Kurtosis
PMV	6.861%	0.4063	8.528%	8.146%	-0.170	1.687
TVM	8.154%	0.5401	10.640%	10.039%	0.363	4.766
PM-EGARCH	6.894%	0.4131	8.685%	8.393%	-0.410	1.647
TVM-EGARCH	7.917%	0.5233	10.146%	9.417%	0.004	4.533
TVM-EGARCHX	8.086%	0.5402	10.398%	9.589%	-0.023	4.149
PQ	6.823%	0.4074	9.036%	9.371%	-0.177	1.672
DQ	7.128%	0.4383	9.272%	9.250%	0.007	3.539
DQX	8.707%	0.6038	11.178%	9.870%	-0.016	6.356

(b) Out-of-Sample Results

	CER	Sharpe Ratio	Mean	St. Dev.	Skewness	Kurtosis
PMV	5.533%	0.2996	8.421%	10.691%	-0.108	2.724
TVM	6.273%	0.3439	8.386%	9.211%	0.296	7.065
PM-EGARCH	5.740%	0.3429	9.175%	11.541%	-0.433	1.590
TVM-EGARCH	7.083%	0.4381	9.603%	10.008%	-0.003	4.035
TVM-EGARCHX	7.042%	0.4320	9.269%	9.378%	-0.114	4.834
PQ	5.392%	0.2968	8.505%	11.076%	-0.175	2.358
DQ	5.036%	0.3000	8.860%	12.138%	-0.448	2.081
DQX	5.603%	0.2951	8.228%	10.199%	0.024	4.992

Note: This table presents summary statistics for the portfolio returns of an investor with power utility and a relative risk aversion coefficient of 5 who uses density forecasts from the models listed in the rows to allocate money to either a risk-free asset or the S&P 500 portfolio. CER is the certainty equivalent return. All values are annualized. All results cover the sample 1956:01-2008:12. The out-of-sample results are obtained by estimating the models recursively using a rolling window with the most recent 30 years of observations. The individual models are described in the caption to Table 3.

Table 7: Risk-Adjusted Relative Return Performance

(a) In-Sample Results

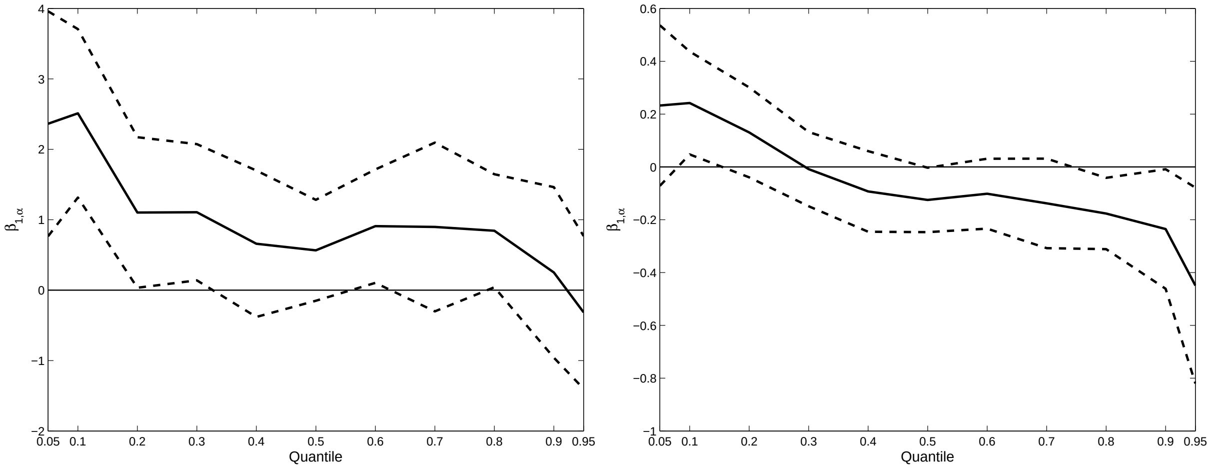
	PMV	TVM	PM-EGARCH	TVM-EGARCH	TVM-EGARCHX	PQ	DQ	DQX
PMV	-							
TVM	2.097%**	-						
PM-EGARCH	0.223%	-1.874%*	-					
TVM-EGARCH	1.713%**	-0.384%	1.490%*	-				
TVM-EGARCHX	2.116%**	0.019%	1.893%*	0.403%	-			
PQ	0.100%	-1.996%**	-0.123%	-1.613%*	-2.016%*	-		
DQ	0.543%*	-1.554%*	0.320%	-1.170%	-1.573%*	0.443%	-	
DQX	2.566%***	0.470%	2.343%***	0.853%	0.450%	2.466%***	2.023%***	-

(b) Out-of-Sample Results

	PMV	TVM	PM-EGARCH	TVM-EGARCH	TVM-EGARCHX	PQ	DQ	DQX
PMV	-							
TVM	1.246%	-						
PM-EGARCH	0.504%	-0.742%	-					
TVM-EGARCH	2.083%*	0.837%	1.578%	-				
TVM-EGARCHX	1.949%*	0.703%*	1.445%	-0.134%	-			
PQ	-0.045%	-1.291%	-0.550%	-2.128%*	-1.995%*	-		
DQ	-0.006%	-1.252%	-0.511%*	-2.089%*	-1.955%*	0.039%	-	
DQX	0.705%	-0.541%	0.201%	-1.378%**	-1.244%**	0.751%	0.711%	-

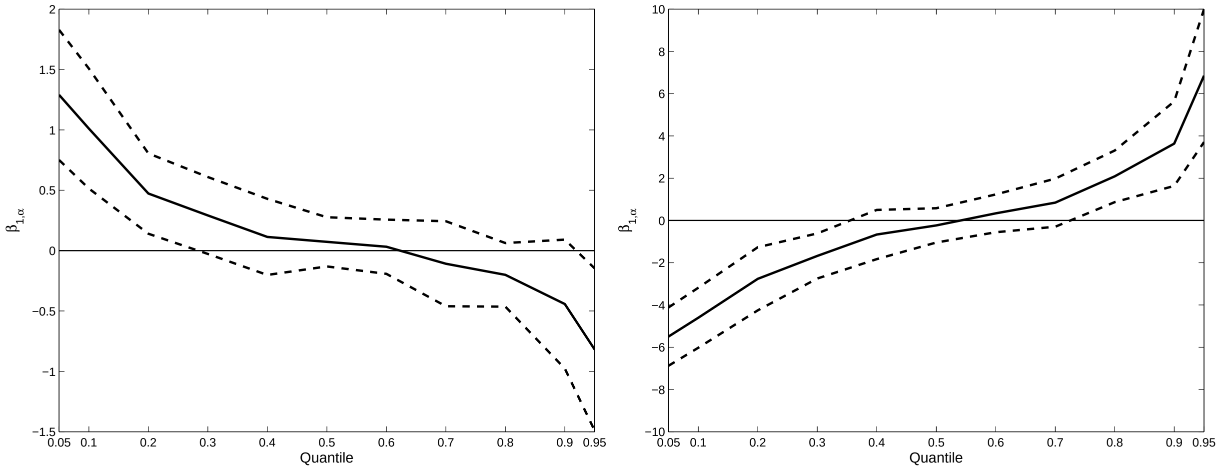
Note: This table reports the relative risk-adjusted return performance (in percent per annum) of portfolios based on models listed in the rows versus those based on models listed in the columns. The risk-adjusted annual alpha is calculated as the sample average of the stochastic discount factor times the difference between portfolio returns based on forecasts from the model listed in the row and those from the portfolio based on the model listed in the column. The stochastic discount factor is based on the approach of Almeida and Garcia (2008). A positive number suggests that the model listed in the row produces higher risk-adjusted returns than the model listed in the column. All results cover the period 1956:01-2008:12. The out-of-sample results are obtained by estimating the models recursively using a rolling window of the most recent 30 years of observations. The individual models are described in the caption to Table 3. * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Figure 1: Slope Coefficients from Linear Quantile Regressions



(a) Dividend Yield

(b) T-Bill Rate

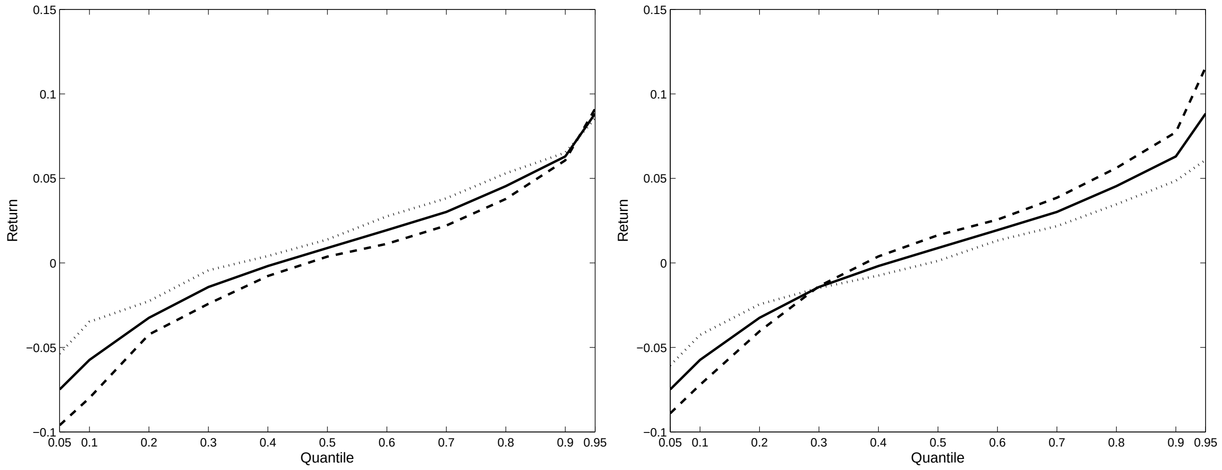


(c) Term Spread

(d) Default Yield

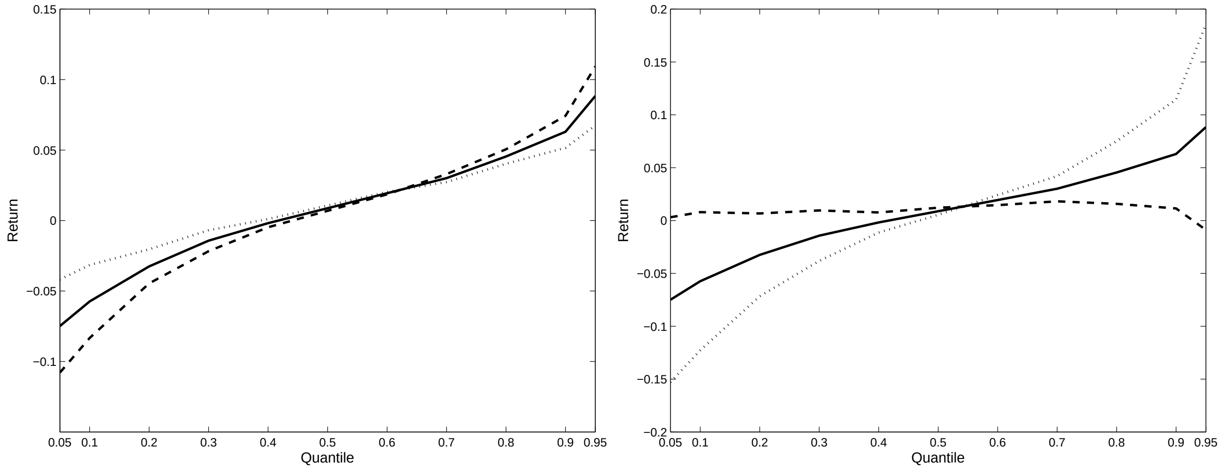
Note: This figure plots the slope coefficients from linear quantile regressions of monthly stock returns on a set of lagged predictor variables. We show estimates of the quantile regression slopes (black solid line) and 95% confidence intervals based on bootstrapped standard errors (black dashed lines).

Figure 2: Quantile Function



(a) Dividend Yield

(b) T-Bill Rate



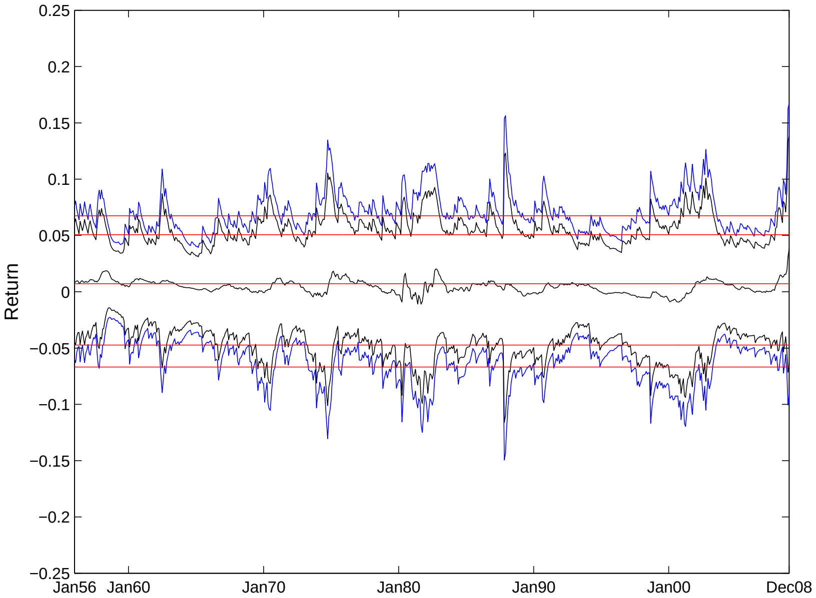
(c) Term Spread

(d) Default Yield

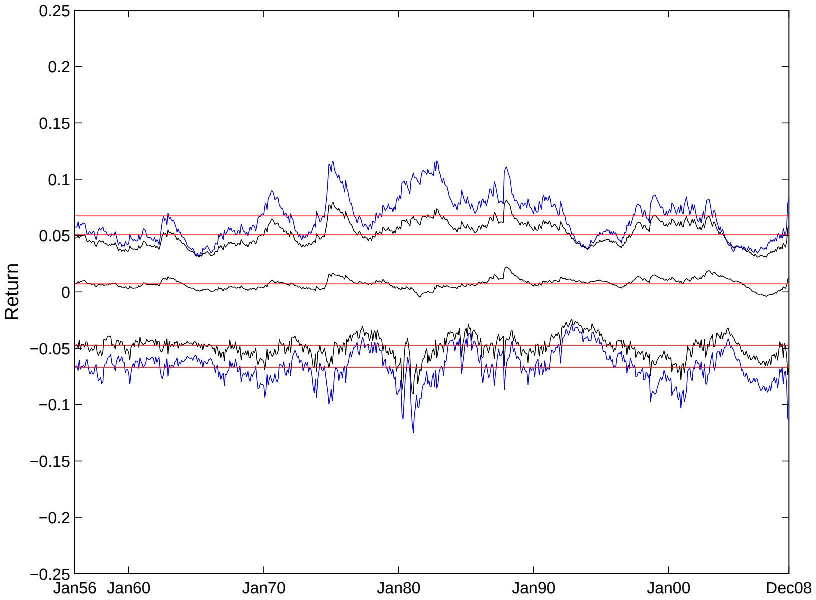
Note: This figure plots the quantile function of returns from the linear quantile model with lagged predictor variables. The solid line sets all predictor variables to their sample mean; the dotted line sets the corresponding predictor variable at its mean plus two standard deviations while the other variables are set to their mean; the dashed line sets the corresponding predictor variable at its mean minus two standard deviations while the other variables are set to their mean.

Figure 3: Time Series of Quantile Forecasts from Dynamic Quantile

(a) TVM-EGARCHX

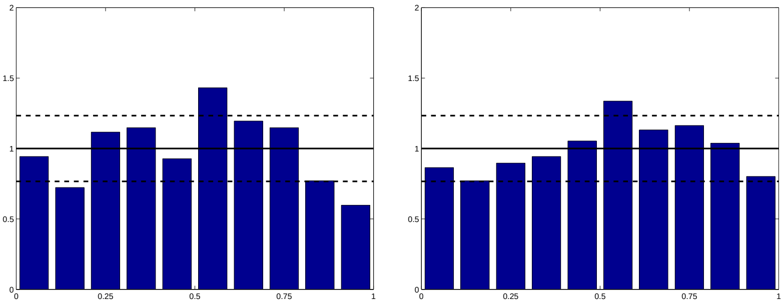


(b) DQX

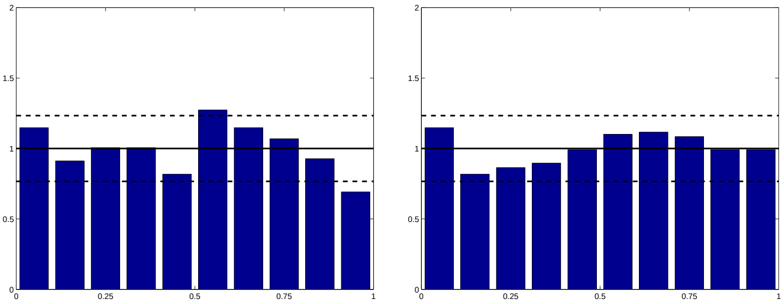


Note: This figure plots time-series of the 5% (bottom blue line), 10% (bottom black line), 50% (middle black line), 90% (top black line) and 95% (top blue line) conditional quantiles from the DQX and TVM-EGARCHX models. The horizontal red lines show the corresponding estimates of the constant quantiles of the return distribution.

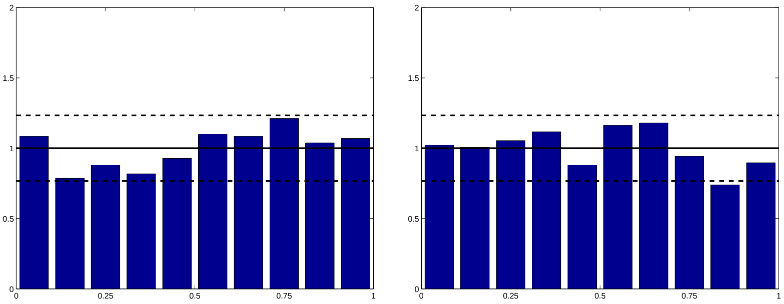
Figure 4: Estimates of the Density of the Probability Integral Transform


(a) PMV

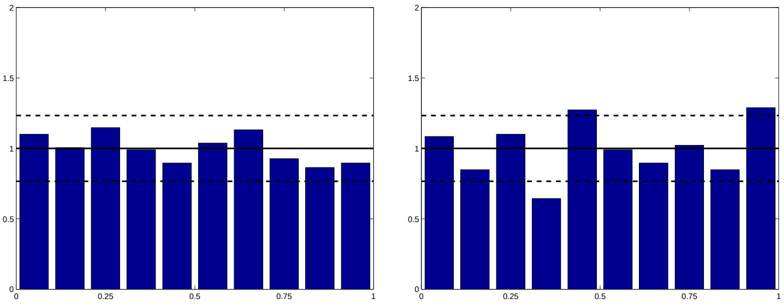
(b) TVM


(c) PM-EGARCH

(d) TVM-EGARCH


(e) TVM-EGARCHX

(f) PQ

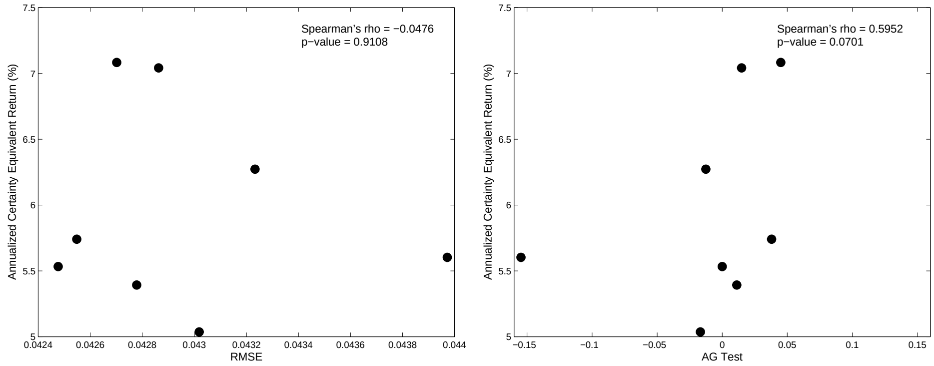

(g) DQ

(h) DQX

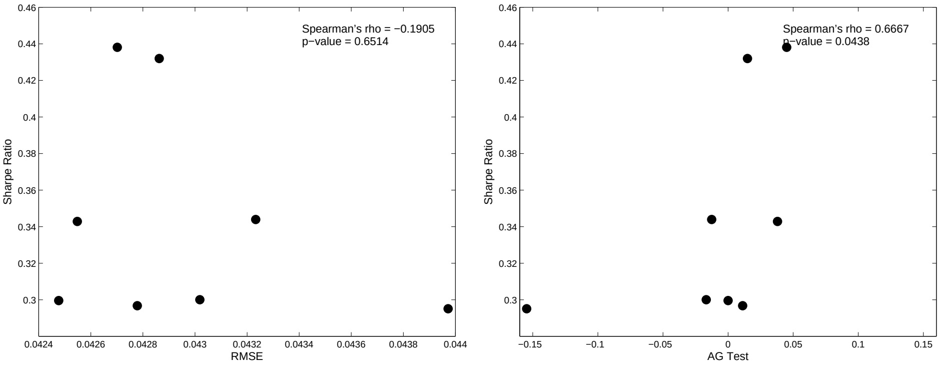
Note: This figure plots histograms of the probability integral transforms evaluated at the actual values of excess returns, u_t, based on out-of-sample return forecasts. The histogram heights are normalized to have an average height of 1. Provided that the density is correctly specified, the sequence of u_t should be independently and identically, uniformly distributed and the normalized histogram should be that of a U(0,1). The solid line represents the probability density function of a U(0,1) random variable while the dashed lines are the 95% confidence intervals for normalized histogram heights. The out-of-sample forecasts for the period 1956:01-2008:12 are obtained by estimating the models recursively using a rolling window with the most recent 30 years of observations.

Figure 5: Rank Correlation between Statistical and Economic Performance Measures

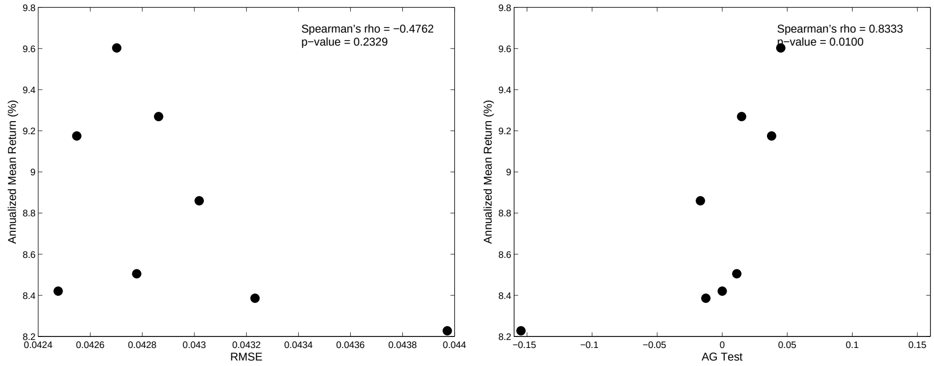
(a) RMSE and CER (b) Prob. Dist. Test and CER



(c) RMSE and Sharpe Ratio (d) Prob. Dist. Test and Sharpe Ratio



(e) RMSE and Mean Return (f) Prob. Dist. Test and Mean Return



Note: This figure presents scatter plots of out-of-sample economic performance measures against statistical performance measures along with the Spearman rank correlation between these measures. RMSE (root mean square error) and probability distribution fit (the Weighted Likelihood Ratio test statistic of Amisano and Giacomini (2007) for the full return distribution) are statistical measures of performance. CER (certainty equivalent return), Sharpe ratio and mean returns are measures of economic performance. All results are computed out-of-sample and cover the period 1956:01-2008:12.