Forecasting the Stock Return Distribution Using Macro-Finance Variables

Yizhen Zhao*

September 18, 2013

Abstract

This paper proposes a new method to forecast $S\&P$ 500 return distribution by combining quantile regression models using macro-finance variables with volatility-based models including various standard EGARCH and stochastic volatility specifications. 30 density forecasting models are compared and combined in an out-of-sample forecasting exercise. Using macro-finance variables is found to help substantially in prediction; the best forecasts are obtained when all 30 models are combined. The proposed density forecasts are shown to be useful to an investor with a CRRA utility function in making optimal portfolio choice. Using the proposed density forecasts yields a certainty equivalent return that is up to 0.35% per month higher than can be obtained with the EGARCH model with a fat-tailed specification.

JEL Classification: C32, C53, E32, E37.

Keywords: Density forecasting, Quantile regression, Portfolio Management

*Corresponding author: Yizhen Zhao, Assistant Professor, East Carolina University, Department of Economics, Greenville, NC 27858, U.S.; Email: zhaoy@ecu.edu; Phone: 410-404-3878; Fax: 252-328-6751. Please do not circulate without permission. I would like to thank my advisors Jonathan Wright, Richard Spady, Laurence Ball, Greg Duffee and Jon Faust for their suggestions and discussion. I also would like to thank Amit Goyal, Ivo Welch, and Jiawei Yao for providing me with the datasets used in this paper. All remaining mistakes are mine.

Contents

1 Introduction

Forecasting the probability distribution of equity return is important for many aspects in finance. Risk managers need to estimate the return distribution on a forward-looking basis to measure the market risk. Portfolio managers may forecast the expected portfolio return using the equity return distribution to make the optimal portfolio choice. Option traders require a precise and timely estimate of the equity return distribution to determine the dynamic hedge ratio. However, it is rare to find a method that is consistently superior - the results can depend on the sample period, the asset class, the frequency of data, and the forecasting horizon. Many macroeconomic policy makers, portfolio managers and private traders have recognized that all models are incomplete descriptions of the reality. As a result, they start to seek answers from forecast combination. Borrowing an idea from the optimal portfolio construction, recent developments in density forecasting combination identify the optimal combination of density forecasts by finding the weights that maximize some indicator of the predictive accuracy. Using such a method, Durham & Geweke (2011) find that combining 42 models with three sources of information about volatility yields a single improved density forecast of daily $S\&P500$ index returns.

This paper is in a similar spirit but adds a new set of forecasts as inputs: quantile density forecasts conditional on the macro-finance variables. Although Macro-finance variables appear to have little value in forecasting the mean of stock returns (Welch & Goyal (2008)), there is some evidence that they can predict other quantiles of the stock return density (Cenesizoglu & Timmermann (2008)). Many of the macro-finance variables considered in the literature are useful in predicting either the left, or right tails of the stock return distribution, but not the entire distribution. In order to obtain the best possible forecast of a future return distribution, a combination of the forecasts using different macro-finance variables might be preferable compared to using a single variable. Accordingly, I take a set of density forecasts, each corresponding to a quantile regression using a single macro-finance variable, and then

find the optimal combination of these forecasts. Furthermore, it would be interesting to know whether combining the density forecasts made by macro-finance variables with those made by volatility-based models can contribute further to density forecasting.

The main contribution of this paper is to find that combining density forecasts made by macro-finance variables with those made by volatility-based models significantly improves density forecasting accuracy. The combined density forecast that incorporates information from both the macro-finance variables and the market volatility, significantly outperforms the density forecast that utilizes either source of information alone. Meanwhile, the combined forecasts using macro-finance variables exhibit better forecasting performance than the combined forecasts using a single source of volatility information. The three sources of volatility considered here are: EGARCH models (the latent volatility), the stochastic volatility, and the realized volatility extracted from the high-frequency data. Adding the implied volatility to the combination exercise impairs the forecasting performance. All procedures are strictly out-of-sample, one-step-ahead predictive distributions for monthly $S\&P500$ index returns from January 1950 to December 2011. The combined predictive densities are constructed in way to maximize the log predictive likelihood following Durham & Geweke (2011). The comparison among forecasts uses the test statistic proposed by Amisano & Giacomini (2007).

The paper also investigates the impact of density forecasts on portfolio decisions. The investor is assumed to have power utility, and needs an estimate of the entire stock return distribution to make optimal portfolio choices. Different density forecasts influence the portfolio choice through different assessments of the expected return using the entire predictive distribution. The portfolio study shows that combined density forecasts that assimilate information from the macro-finance variables yield a certainty equivalent return that is up to 0.35% per month higher than can be obtained with the EGARCH model with a fat-tailed specification. The advantage of using macro-finance variables are tested through a few turmoil periods of the stock market. Portfolios that use combined density forecasts with macro-finance variables can significantly better capture the market surge. Furthermore, as

investors become more risk averse, the difference in gain brought by different density forecasts narrows, and the portfolio choices made by users of different forecasts converge.

The remainder of the paper proceeds in four sections. Section 2 lays out the five classes of density forecasting models and the density combination technique. Section 3 presents a real-time portfolio study that examines the impact of the combined density forecasts introduced in section 2 on optimal portfolio choices. Section 4 concludes.

2 Models and Estimation

2.1 Forecasting Models

There are five classes of models that are being considered and compared: (I) combined quantile density forecasting models using macro-finance variables, (II) single-model quantile density forecast using macro-finance variables or their principal components, (III) **EGARCH** models, (IV) stochastic volatility models (**SVOLs**, or **SVs** for short), and (V) realized volatility models using high-frequency data (**RVs**). Each class contains four to eight models. Each model provides an approximation to the stock return distribution but none of them is literally true. Models within each class or across classes are combined to form the predictive densities at the end of each month. The premise of density forecast combination is to combine multiple data generating processes and various sources of information. The forecaster is assumed to stand at the close of trading month $t-1$ and tries to make density forecasts for stock return at month t.

Quantile Density Forecasting This approach involves modeling a particular quantile of the stock return distribution using macro-finance variables. Intuitively, if a macro-finance variable consistently well predicts the lower quantiles ($\tau < 10\%$), then it is possible that

the variable can predict the market crash. If some variable always exhibits strong power in predicting only the upper quantiles ($\tau > 90\%$), its movement may imply a market surge. Specifically, let $Q_\tau(y_t|x_{t-1})$ be the value such that τ (percent) of the mass of the distribution $F(y_t|x_{t-1})$ is less than $Q_\tau(y_t|x_{t-1})$. Each quantile is predicted by

$$\hat{Q}_\tau(y_t|x_{i,t-1}) = \hat{\beta}_{i0}(\tau) + \hat{\beta}_{i1}(\tau)y_{t-1} + \hat{\beta}_{i2}(\tau)x_{i,t-1}. \quad \text{for } i = 1\ldots, N, \tag{1}$$

where i refers to each of the N macro-finance variables. $\hat{\beta}_{i0}(\tau)$ and $\hat{\beta}_{i1}(\tau)$ are estimated following Koenker & Bassett (1978). They describe how the τ-th quantile of the future return might vary with a particular variable, $x_{i,t-1}$. By choosing a fine grid of quantiles, say, τ's: $0.01, \ldots, 0.99$, one can trace out the entire return distribution conditional on $x_{i,t-1}$. Since the estimation uses finite sample, the estimated quantiles might be too close to one another, or even the same, and may exhibit "crossings"[1], which causes the density estimation that directly flips the difference of quantiles unreliable. Alternatively, the predictive density of stock return at t can be estimated by the nonparametric kernel density:

$$f(y_t|x_{i,t-1}) = \frac{1}{nh_n}\sum_{\tau=0.01}^{0.99} K\big(\frac{y_t - \hat{Q}_\tau(y_t|x_{i,t-1})}{h_n}\big), \quad \text{for } i = 1, \ldots, N.$$

This approach was also used by Gaglianone & Lima (2009) and Zhao (2011). Moreover, each macro-finance variable exhibits different predictive power on different quantiles (Cenesizoglu & Timmermann (2008)). Different models of macro-finance variables may give different density forecasts. Weighting these densities in way that maximizes a certain measure of the predictive accuracy may give a relatively precise shape of the future return distribution.

Another recent development in the forecasting literature has been the recognition that using the first few principal components from a large dataset can avoid estimation of too many parameters, and is helpful for *point forecasting*. (see Stock & Watson (1998), Stock

[1]To deal with this issue, I follow Chernozhukov et al. (2010)'s approach and rearrange the original quantile estimates into ascending order.

& Watson (2003), Stock & Watson (2005)). Zhao (2011) and Manzan & Zerom (2009) find that using the first few principal components from a large dataset is also helpful in *density forecasting*. I hence considered **density forecasts using the first few principal components** of all N macro-finance variables. In particular, each quantile of y_t is predicted by

$$\hat{Q}_\tau(y_t|\mathbf{x}_{t-1}) = \hat{\beta}_0(\tau) + \hat{\beta}_1(\tau)y_{t-1} + \hat{\beta}_2(\tau)\mathbf{f}_{t-1}. \tag{2}$$

where $\mathbf{f}_{t-1}$ are the first $\mathbf{r}$ principal components of $\underbrace{\mathbf{x}_{t-1}}_{N\times 1}$. $\mathbf{r}$ is set to be 1 in the empirical study[2]. In addition, to see whether it is better to use different number of factors in forecasting different quantiles, the quantile-varying factor selection rule (**ATIC**) proposed by Ando & Tsay (2011) are taken into account. The details of the ATIC factor selection rule are enclosed in Appendix C.

Furthermore, I also considered a **multivariate density forecast** - forecasting each quantile using all N macro-finance variables directly as conditioning variables. It simply replaces $\mathbf{f}_{t-1}$ in equation (2) by $\underbrace{\mathbf{x}_{t-1}}_{N\times 1}$. Although the forecast also uses the information extracted from macro-finance variables, its forecasting performance is significantly worse (see Section 2.3). Forecast combination plays a critical role in improving the forecasting performance. In reality, it is highly likely that the dynamics of the stock return is driven by multiple sources of information, while the true data generating process is a combination of multiple distributions.

EGARCH Following Bollerslev (1987), the EGARCH model takes the form:

$$y_t = \mu_Y + \sigma_Y \exp(\sum_{i=1}^{k} h_{i,t}/2)\varepsilon_{j,t} \tag{3}$$

$$h_{i,t} = \alpha_i h_{i,t-1} + \beta_i(|\varepsilon_{j,t-1}| - (2/\pi)^{1/2}) + \gamma_i \varepsilon_{j,t-1}. \tag{4}$$

[2]When $\mathbf{r} > 1$, the forecast accuracy is worse than when $\mathbf{r} = 1$.

where y and h are the vectors of data and volatilities. $h_{i,t}$ has up to two components (i = 1, 2) - one captures a persistent long-term trend in the level of volatility; the other captures short-term fluctuations around it.

The innovations in the conditional mean equation, ε_t, is specified to distribute as Gaussian, Student's t- (Bollerslev (1987)) and the Generalized Error Distribution (GED) (Nelson (1991)). Gaussian ε_t has up to two normal components (j = 1, 2). The two fat-tailed distributions capture the excess kurtosis of stock return. ε_t is always standardized to have mean zero and variance one.

Parameters μ_Y, σ_Y, α_i, β_i, and γ_i are estimated by maximum likelihood estimation. If $\gamma_i < 0$, the model captures the *leverage effect*, in which volatility tends to rise in response to "bad news" and to fall in response to "good news".

Since the next period's volatility in EGARCH has already been determined when the forecast is made, y_t is a monotonic function of ε_t. The predictive distribution of y_t thus has closed form. According to the density transformation theorem, the density of y_t is just a linear function of that of ε_t. In particular, $f_Y(y) = f_\varepsilon(g^{-1}(y))|\frac{\partial g^{-1}(y)}{\partial y}|$, and $g^{-1}(y) = \varepsilon = (y - \mu_Y)/[\sigma_Y \exp(\sum_{i=1}^{k} h_{it}/2)]$. Eight models estimated in real-time and later used to form the combined density forecasts in this class are: $EGARCH - Gaussian_{i=1,j=1}$, $EGARCH - Gaussian_{i=1,j=2}$, $EGARCH - Gaussian_{i=2,j=1}$, $EGARCH - Gaussian_{i=2,j=2}$, $EGARCH - t_{i=1,j=1}$, $EGARCH - t_{i=2,j=1}$, $EGARCH - GED_{i=1,j=1}$ and $EGARCH - GED_{i=2,j=1}$.

Stochastic Volatility Model Stochastic volatility models offer a natural alternative to the GARCH family of time-varying volatility models. The general form of the univariate

SVOL model[3] can be represented by

$$
\begin{aligned}
y_t &= \exp(h_t/2)\sqrt{\lambda_t}\varepsilon_t, && (5)\\
h_t &= \mu + \phi(h_{t-1} - \mu) + \eta_t, \qquad t = 1, \ldots, T. && (6)\\
\eta_t &= \rho\varepsilon_t + \sqrt{1-\rho^2}u_t, \qquad u_t \sim N(0,1). && (7)\\
\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} &\mid (\rho, \sigma) \sim i.i.d.\mathcal{N}_2(0, \Sigma), && (8)\\
\Sigma &= \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix}. && (9)
\end{aligned}
$$

The parameters are $\theta = (\phi, \sigma, \rho)$ and μ, where μ is the intercept, ϕ is the volatility persistence and σ is the standard deviation of the shock to h_t. The basic univariate stochastic volatility (SVOL) model specifies that conditional volatility follows a log-normal auto-regressive process ($\lambda_t = 1$) with innovations assumed to be independent of the innovations in the conditional mean equation ($\rho = 0$) (see Kim et al. (1998)).

Since equity return data commonly exhibit volatility clustering and non-Gaussian distributions, the basic SVOL model can be extended in two ways: (1) with a fat-tailed distribution of the conditional mean innovations and (2) with a *leverage effect* via correlation between the volatility and the mean innovations (see Jacquier et al. (2004) and Omori et al. (2007)).

The fat-tailed SVOL model specifies $\nu/\lambda_t \sim \chi^2_\nu$, in which case, $\sqrt{\lambda_t}\varepsilon_t$ follows a student-t distribution with ν degree of freedom, $\nu \sim Gamma(16, 0.8)$, and $\lambda_t^{-1} \sim Gamma(\nu/2, \nu/2)$. The fat-tailed SVOL can deal with outliers by introducing a large λ_t without increasing h_t. It is more outlier resistant than the basic SVOL.

Meanwhile, the correlated SVOL model specifies the innovations of the volatility to be correlated with the innovations of the conditional mean. Negative correlations ($\rho < 0$) between mean and variance errors can produce a *leverage effect* in which negative (positive)

[3]referred as SV in tables and figures.

shocks to the mean are associated with increases (decreases) in volatility. For example, if $\rho = -0.6$ and $|\varepsilon_t| = 1.5$, then the expected volatility will be 60% higher for negative versus positive shocks. The asymmetry in the correlated SVOL model also induces left-skewness in the marginal distribution of the stock return, which is consistent with the non-parametric evidence of Gallant et al. (1997).

To estimate SVOL, each model is viewed as a hierarchical structure of three conditional distributions: the conditional distribution of stock returns, $p(\mathbf{y}|\mathbf{h})$; the conditional distribution of volatility, $p(\mathbf{h}|\theta)$; and the marginal or prior distribution of parameters, $p(\theta)$. Estimation of SVOL models uses the Markov-Chain Monte Carlo method. The choice of priors and posteriors of the parameters follows Kim et al. (1998), Jacquier et al. (2004) and Omori et al. (2007). $\mathbf{h}|\theta$ is smoothed by augmented Kalman filter as in Kim et al. (1998) and Omori et al. (2007). Estimation details are enclosed in the Appendix A.

In sum, as the actual realizations of return volatility are not directly observable, EGARCH and SVOL models deal with the fundamental latency of return volatility through strong parametric assumptions. An alternative approach is to invoke option pricing models to invert observed equity prices into market-based forecasts of implied volatility over a fixed future horizon. However, these procedures are model-dependent and incorporate a volatility risk premium in the measure, so that they generally do not provide unbiased forecasts of the volatility of the underlying asset. This is justified by four models in the subsection of **Implied Volatility Model**.

Realized Volatility Model Realized volatility is a nonparametric ex-post estimate of the return variation. The most obvious realized volatility measure is the sum of finely-sampled squared return realizations over a fixed time interval. In principle, high-frequency returns are capable of providing very precise information about the latent volatility state. Volatility

filtered from high-frequency data is often used as a proxy for the variance in that month. This uses the insight of Merton (1980) and Nelson (1992) that volatility may be arbitrarily precisely estimated using sufficient high-frequency data in a frictionless market.

Since the macroeconomic variables are collected at the monthly frequency, this study uses the sum of the daily squared returns to approximate the monthly variance. Return data at the higher frequency is not used because they are likely affected by various market microstructure frictions or noise, arising from bid-ask bounces, a discrete price grid, and so on. In particular, the monthly variance is calculated using the equation.

$$\sigma_t^2 = \sum_{i=1}^{N_t} r_{it}^2 + 2 \sum_{i=2}^{N_t} r_{it} r_{i-1t}.$$

where N_t is the number of trading days in month t and r_{it} is the return on the ith day of month t. The second term accounts for the autocorrelation observed in daily returns. Using this correction introduced by French & Stambaugh (1987) slightly improves the forecast accuracy. σ_t^2 is then treated as the realized volatility (RV) in modeling the return:

$$y_{t+1} = \mu_t + \sigma_t \varepsilon_t, \tag{10}$$

$$\log \sigma_t^2 = \phi_0 + \phi_1 \log \sigma_{t-1}^2 + u_t. \tag{11}$$

The logarithm of the realized volatility is assumed to follow an AR(1). ϕ_0, and ϕ_1 are estimated via OLS. The specification of the return innovations ε_t is the same as those in the EGARCH models. μ_t is estimated by maximum likelihood. The predictive density of y_t is formed by integrating across uncertainty in the volatility state $\hat{\sigma}_t \varepsilon_t$. Four models considered in this class include: $RV - Gaussian_{j=1}$, $RV - Gaussian_{j=2}$, $RV - t$, and $RV - GED$. Empirical results show that the density forecasting accuracy of this class of model critically depends on the distribution chosen for the return innovations.

Implied Volatility Model In addition, option-pricing models allow us to "back out" market estimates of stock-price volatility, which is referred to as the implied volatility. Based on real-time option prices, the implied volatility reflects investors' consensus view of future expected stock market volatility. The Chicago Board Options Exchange (CBOE) regularly computes the implied volatility of $S\&P$ 500 (VIX) since January, 1990, and the implied volatility of $S\&P$ 100 (VXO) since January, 1986. Both indices measure the market's expectation of future 30-day volatility of $S\&P$ 100 index and $S\&P$ 500 index respectively.

To exploit the information contained in the implied volatility, I considered two cases. The first is to model the return as in equation (10), but substitute σ_t with the end-of-month VIX or VXO. Due to the four different specifications of the return innovation, four models of each implied volatility index are combined with other 30 forecasts. The combination starts from January, 1986 for VXO, and from January, 1990 for VIX, with all other 30 forecasts using information up to January, 1950. The second method uses the implied volatility as one of the conditioning variables to perform the quantile density forecast as in equation (1), the forecast of which is then combined with forecasts that use N macro-finance variables. The combination starts from January, 1995. It turns out that the implied volatility does not help improving the predictive accuracy in either case. The failure of the implied volatility might be due to that it is not an unbiased estimator of the return volatility.

2.2 Forecast Combination and Comparison

All forecasts are out-of-sample and in real-time. I adopt the recursive method to obtain the out-of-sample forecasts, in which the parameters are updated as the forecast moves forward through time, with the forecasting origin fixed to be the first in-sample period. The predictive accuracy of any density forecast is measured by the *log predictive likelihood*, which is the sum of logarithmic predictive densities over the out-of-sample forecasting periods.

At the end of month $t-1$, $\{f(y_s|\mathbf{x}_{s-1}, \mathbf{y}_{s-1}, A_m)\}_{s=q+1}^{t-1}$ is the historical predictive density of model A_m, in which q is the number of observations used to construct the first forecast. The forecaster combines the models in each class by finding the weights that maximize the log predictive likelihood of that class. In particular, the optimal weight vector $\mathbf{w}_{t-1}^*$ is chosen to maximize:

$$f_{t-1}(\mathbf{w}_{t-1}) = \sum_{s=q+1}^{t-1} \log\big[\sum_{m=1}^{M} w_{t-1,m} f(y_s|\mathbf{x}_{s-1}, \mathbf{y}_{s-1}, A_m)\big]. \tag{12}$$

M is the number of models that are being combined. $\mathbf{w}_{t-1}^* = (w_{t-1,1}^*, \ldots, w_{t-1,M}^*)'$ is a weight vector satisfying

$$\sum_{m=1}^{M} w_{t-1,m}^* = 1, \qquad w_{t-1,m}^* \geq 0, \qquad \text{for } m = 1, \ldots, M.$$

Given any interval on the support of the stock return distribution, different forecasts may result in different predictive densities. The optimal weight $\mathbf{w}_{t-1}^*$ is then used to weight these predictive densities to form an estimate of the true density corresponding to this particular interval. The predictive density of y_t is the weighted average of the predictive densities obtained by each model by the end of $t-1$:

$$f(y_t|\mathbf{x}_{t-1}, \mathbf{y}_{t-1}, \mathbf{w}_{t-1}^*) = \sum_{m=1}^{M} w_{t-1,m}^* f(y_t|\mathbf{x}_{t-1}, \mathbf{y}_{t-1}, A_m). \tag{13}$$

Geweke & Amisano (2011) show that $f_{t-1}(\mathbf{w}_{t-1})$ is at least weakly concave, and for $t > M$, $f_{t-1}(\mathbf{w}_{t-1})$ is in general strictly concave. Given the concavity of the log likelihood function, the optimal pool[4] can (and often does) assign positive weight to all participating models.

The formation of optimal prediction pool has interesting parallels to optimal portfolio

[4]Any prediction rule combines many individual prediction rules with weights nonnegative and summed to one, is referred as a prediction pool.

construction without short sales. Given a collection of risky assets, the optimal portfolio will typically include a mix of risky assets rather than placing all weight on the single asset with the highest Sharpe ratio. Even though a particular asset may have a comparatively low return when it is negatively correlated with the market return, it can improve the performance of the portfolio through diversification. The optimal portfolio will typically have better performance than any of the individual assets alone.

These same points also hold true for the optimal prediction pool. Given a collection of models, the goal is to maximize log predictive likelihood. The optimal pool typically includes a mix of models rather than placing all weight on the single model with the highest predictive likelihood. Even though a particular model may produce an inferior predictive likelihood on average, it can enter the optimal prediction pool with positive weight if it occasionally but regularly outperforms the other prediction models. The pool will generally have better performance than any of the models it comprises.

The optimal prediction pool fundamentally differs from the Bayesian model averaging and the conventional forecast competition. The Bayesian model averaging identifies the true model as sample size grows, by assigning all weight to the model with the highest expected predictive likelihood under the data generating process. The conventional forecast competition identifies the model closest to the data generating process in Kullback-Leibler distance asymptotically, which leads to a horse race with a single ultimate winner. Nonetheless, in the optimal prediction pool, each model contributes a strength that balances some weakness of the other models. The optimal prediction pool proves superior to the single-model forecasts and wins the horse race in the conventional forecast competition.

Figure.1 illustrates how the individual density forecasts are combined to form a predictive distribution. The forecast is made one-month ahead for June, 2008. The blue line represents the predictive distribution made by book-to-market ratio. The red line represents the predictive distribution made by term-spread. The black line is the combined predictive

distribution produced by the optimal predictive pool using all 11 macro-finance variables[5].

To compare of the out-of-sample forecasting performance of any two forecasts, I use the likelihood ratio test developed by Amisano & Giacomini (2007). The test statistic takes the form of a t-statistic:

$$AG_{q,T} \equiv \frac{\Delta\overline{L}_t(y_{t+1})}{\hat{\sigma}/\sqrt{(T-q)}},$$

where

$$\Delta\overline{L}_t(y_{t+1}) \equiv \frac{1}{T-q}\sum_{t=q+1}^{T} L_t(y_{t+1}) = \frac{1}{T-q}\sum_{t=q+1}^{T} \log f(y_{t+1}|\mathbf{x}_t,\mathbf{y}_t) - \log g(y_{t+1}|\mathbf{x}_t,\mathbf{y}_t).$$

$f(\cdot)$ and $g(\cdot)$ are the predictive densities of two competing density forecasts. $\hat{\sigma}^2$ is a finite-sample estimate of the asymptotic variance of $\Delta L_t(y_{t+1})$ [6]. The null and the alternative are stated as

$$H_0 : E[\Delta L_t(y_{t+1})] = 0 \;\; v.s \;\; H_A : E[\Delta L_t(y_{t+1})] > 0.$$

The null hypothesis will be rejected if model-$f(\cdot)$ provides more accurate density forecasts relatively to model-$g(\cdot)$, in which case, $AG_{q,T} > 0$. When the models being compared are nonnested, the asymptotic distribution of the test statistic follows the standard normal distribution. When a combined forecast is compared with a forecast that it incorporates,

[5]The optimal pooling weights assigned to the 11 macro-finance variables are: dividends**(D12)**(0.0000), earnings**(E12)**(0.0061), stock variance**(svar)**(0.0000), book to market ratio**(b/m)**(0.2325), net equity expansion**(ntis)**(0.0000), term spread**(tms)**(0.3048), default yield spread**(dfy)**(0.0000), inflation**(infl)**(0.1711), unemployment rate**(ume)**(0.0297), industrial production growth**(ip)**(0.2557), non-farm payroll**(nfp)**(0.0000).

[6]For $h > 1$, $\hat{\sigma}^2$ uses the heteroskedasticity and autocorrelation consistent (HAC) estimator of the asymptotic variance(see Newey & West (1987)). For one-period ahead forecast ($h = 1$), the choice of lag truncation is set to be 0 as in Giacomini & White (2006) and Amisano & Giacomini (2007).

the forecasts that are compared might be nested[7], in which case, the asymptotic distribution of the AG statistic is not normal[8].

To check if there is distortion in the distribution of the statistic, bootstrap *p-value*s are reported. The procedure starts with re-sampling from $\{\begin{bmatrix} f(\cdot) & g(\cdot) \end{bmatrix}\}_{t=q+1}^{T}$ for $N = 500$ times and produces N bootstrap test statistics: $AG_{q,T}^{boot}$. When there is no distortion, the bootstrap test statistics form a distribution that is the same as the asymptotic distribution. The $1-\alpha$ confidence interval (one-sided) is then constructed using the percentile-t interval, as in

$$\{AG_{q,T} : AG_{q,T} \leq \bar{AG}_{q,T}^{boot} - \hat{\sigma}_{AG}\tilde{F}_{\alpha}^{boot}\},$$

where $\bar{AG}_{q,T}^{boot}$ and $\hat{\sigma}_{AG}$ are the mean and the standard deviation of the 500 bootstrap test statistics. $\tilde{F}^{boot}$ refers to the bootstrap distribution of the AG test statistics.

Furthermore, one may compare the predictive accuracy of competing density forecasts over a specific region of interest. In particular, The predictive accuracy given a region of interest can be measured by the censored likelihood(csl) score function (see Diks et al. (2011)), given by

$$S^{csl}(\hat{f}_t; y_{t+1}) = I(y_{t+1} \in A_{t+1}) \cdot \log \hat{f}_t(y_{t+1}) + (1 - I(y_{t+1} \in A_{t+1})) \cdot \log\big(1 - \int_{A_{t+1}} \hat{f}_t(\mathbf{y}) d\mathbf{y}\big), \quad (14)$$

where A_{t+1} is the region of interest, and $\hat{f}_t(y_{t+1})$ is estimated by $f(y_{t+1}|\mathbf{x}_t, \mathbf{y}_t)$. This scoring rule takes into account the accuracy of the density forecast for the total probability of y_{t+1}

[7]Two models are nested if one is a special case of the other; they are nonnested if neither can be represented as a special case of the other. Vuong (1989) provided a general distribution theory for the likelihood ratio test that covers nested and nonnested models, which tests whether the two densities f and g have the same Kullback-Liebler information criterion. The true data generating process can be unknown and differ from both f and g.

[8]The issue is familiar in the context of point forecasting. For point forecasting, the analog of the AG statistic is the test proposed by Diebold & Mariano (1995). When *point forecasts* are nested, Diebold & Mariano (1995) test statistic has a nonstandard distribution, derived by Clark & McCracken (2001), Clark & McCracken (2005), McCracken (2007), which is a function of stochastic integrals of Brownian motion. When *density forecasts* are nested, no corresponding results are known.

falling into the region of interest. The test statistic takes the same form as Amisano & Giacomini (2007) test above.

2.3 Forecasting Results

In this section, density forecasts made by individual models are first grouped and combined according to the source of information they use. Then these combined density forecasts are compared with one another. Afterwards, all 30 models are combined together to generate a single combined density forecast that incorporates the information from both macroeconomic variables and the market volatility.

I take the sample period from January, 1950 to December, 2011. The stock return series is derived from the $S\&P500$ price index data, p_t, as $y_t = 100\log(p_t/p_{t-1})$, so that y_t represents the continuously compounded return on the index. I use eleven macro-finance variables to make the forecast - dividends**(D12)**, earnings**(E12)**, stock variance**(svar)**, book to market ratio**(b/m)**, net equity expansion**(ntis)**, term spread**(tms)**, default yield spread**(dfy)**, inflation**(infl)**, unemployment rate**(ume)**, industrial production growth**(ip)**, and non-farm payroll**(nfp)**. The first eight variables are taken from the dataset maintained by Goyal and Welch[9]. The last three variables are taken from the Federal Reserve Bank of Philadelphia's Real-Time Data Set for Macroeconomists (RTDSM). The data set reflects what the macroeconomic data exactly looked like at each historical date, so that forecasting using the dataset unveils what the forecast would have performed as they looked at the time (Clark & McCracken (2009)). As for realized volatility models, the daily return data from January 2, 1950 through December 31, 2011 are obtained from CRSP value-weighted returns (including dividends).

[9]I pick the sample period also because December, 2011 is the last observation available in Goyal and Welch's dataset by October, 2012. The Additional details on data sources and the construction of these variables are provided by Welch & Goyal (2008).

Table.1. reports the log predictive likelihood obtained by the above 30 density forecasts. The higher the predictive likelihood, the better is the forecast. For legibility, I use the EGARCH Student's-t($i = 1, j = 1$) as the benchmark forecast. The reported value is the difference in the log predictive likelihood between the forecast of interest and the EGARCH Student's-t($i = 1, j = 1$), which has a log predictive likelihood of -1858.2. Besides, $\Delta(A_i, A_j) = \exp\{[\sum_{t=q+1}^{T} L_t(y_{t+1}, A_i) - \sum_{t=q+1}^{T} L_t(y_{t+1}, A_j)]/(T-q)\}$ represents a geometric average proportional difference in predictive densities. A difference of 6.4080 in log predictive likelihood generally corresponds to a 1% increment in "probability" over the entire sample period.

The first column reports the predictive likelihood achieved by combined density forecasts. The other columns report that of all single-model forecasts. The combined density forecast using all 30 models ($Comb.All$) exhibits the highest predictive likelihood. The combined density forecast using the realized volatility ($Comb.RV$) is the second best. The combined quantile density forecast using all N variables ($Comb.MF$) and the single-model forecast with ATIC-selected factors ($ATIC - Factor$) outperform the combined forecast of EGARCH models ($Comb.EGARCH$) and the combined forecast of stochastic volatility models ($Comb.SV$). Even using only the first eight macro-finance variables, $Comb.MF$ and $ATIC - Factor$ have outperformed many individual models that use volatility information alone, though not the combined forecasts. Adding the last three real-time macroeconomic variables substantially improves the predictive accuracy. The success of $Comb.MF$ and $ATIC - factor$ owes to utilizing different conditioning information in forecasting the different parts of the return distribution. In addition, adding implied volatility to the combination exercise impairs the forecasting performance. Combining all 30 models with VXO index reduces the log-predictive likelihood by 2.55. Combining all 30 models with VIX index reduces the log-predictive likelihood by 2.76. Using implied volatility in quantile density forecasting leads to a log-predictive likelihood 262.05 points lower than the benchmark forecast.

Table.2.*a* reports the results of the Amisano & Giacomini (2007) test. The row of the table

represents the forecast that is being compared to, as $f(\cdot)$ in constructing the test statistic. The column of the table represents the forecast that is used to compare $f(\cdot)$, as $g(\cdot)$ in constructing the test statistic. Given that all combined forecasts are mixtures of multiple data generating processes, their major difference is driven by the source of information they use. The comparison between two forecasts in fact compares the predictive power of different conditioning information. As a complement, $Table.2.b$ reports the test results when comparing single-model forecast ($g(\cdot)$) to the combined forecast using all models ($f(\cdot)$). A positive significant AG-test statistic shows that the combined density forecast of all 30 models significantly outperform any single-model forecast in above five classes. Two main findings are summarized below.

First, the combined forecast using all 30 **models does best.** The combined density forecast using various sources of information significantly outperforms combined forecasts that use a single source of information. When the comparison is set between combined forecast using all models ($Comb.All$) and any other forecasts except combined forecast using realized volatility ($Comb.RV$), the AG statistics are all positive and significant at 5%. When the comparison is set between $Comb.All$ and $Comb.RV$, the AG statistic remains positive. The significance depends on the sample period. When the sample ends at December, 2008, the AG statistic is significant at 10% level. The significance holds when the inference uses either bootstrap critical values or the asymptotic critical values.

Second, the combined density forecasts that use macro-finance variables exhibit higher forecasting accuracy than the combined forecasts of volatility-based models. Over the entire 624 out-of-sample periods, combining the quantile density forecasts of all 11 macro-finance variables outperforms the combined forecast of stochastic volatility models at the 10% level of significance; and outperforms the combined forecast of EGARCH models at the 5% level of significance. Using the same set of conditioning information but without forecasts combination, quantile density forecasts using the principle components extracted from macro-finance variables, including $ATIC-Factor$ and $1st.PC$, exhibit compet-

itive predictive accuracy with the combined forecast of EGARCH models ($Comb.EGARCH$) and the combined forecast of stochastic volatility models ($Comb.SV$)[10]. In contrast, quantile density forecast that uses the same set of macro-finance variables in predicting all different quantiles is significantly worse than $Comb.EGARCH$ and $Comb.SV$. The success of $Comb.MF$ and $ATIC - Factor$ implies using quantile-varying information extracted from the macroeconomic variables is more powerful than using the same set of conditioning variables for the entire support of the return distribution.

In addition, $Figure.2.$ displays how the optimal pooling weights $\{w^*_{t,m}\}^T_{t=q+1}$ are allocated among different types of forecasts. The horizontal axis represents the sum of the weights assigned to the forecasts made by macro-finance variables. The vertical axis represents the sum of the weights assigned to the forecasts made by volatility information only. Each contour line centers at the point that shows how the optimal pooling weights are allocated between the two types of forecasts, and represents an out-of-sample forecasting period. The most dense region, where the contour lines cluster, locates at $(0.4237, 0.5763)$, which indicates that the macro-finance variables contribute 40% forecasting accuracy to the optimal pool, while the volatility models contribute 60% forecasting accuracy to the optimal pool.

2.4 A Comparison with the State Price Density

The *state price density* (SPD) is the continuous-state counterpart to the price of Arrow-Debreu securities each paying one dollar in one specific state of nature and nothing in any other state. The SPD aggregates important information regarding investors' preferences, behavior, and expectations of the market, and is widely used for pricing and hedging. In the Black-Scholes model, SPDs are log-normal distributions with constant volatility. In practice, the volatility is time-varying and the stock price deviates from log-normal distribution. The

[10]The AG test statistic is positive but insignificant.

state price density is the second derivative (normalized to integrate to unity) of a call option pricing formula with respect to the strike price (see Ross (1976), Banz & Miller (1978) and Breeden & Litzenberger (1978)).

I considered two methods to obtain the state price density. **Method I** forms a money spread with two call options, and assuming the probability that the stock price falls into the interval between the strike prices of these call options to be zero. Consider the portfolio obtained by buying one call options struck at X_j and selling another call option with identical expiration date, but at higher exercise price X_{j+1}. This portfolio pays 0 if $S_T \leq X_j$, and $X_{j+1} - X_j$ if $S_T > X_{j+1}$. Now denote by $H(S_t, X, \tau)$ the market price of a call option at time t with strike price X, time-to-maturity τ, and the underlying asset price S_t. The price of the money spread must be

$$H(S_t, X_j, \tau) - H(S_t, X_{j+1}, \tau) = 0 \cdot P(S_T \leq X_j) + (X_{j+1} - X_j) \cdot P(S_T > X_{j+1}).$$

Using a fine grid of strike prices $j = 1, \ldots, n$, $P(S_T \leq X_j) = 1 - P(S_T > X_{j+1})$ yields the cumulative distribution function of the future stock price S_T. The state price density f^* is obtained by taking the derivative of the cumulative distribution function with respect to the strike price. Since the $S\&P$ 500 index options expire on Saturdays following the third Friday of the expiration month, one-month ahead forecast is made using the call option prices on the third Friday of the current month with the days to maturity restricted to be 30.

Following Aıt-Sahalia & Lo (1998), **Method II** takes the option pricing formula H to be a nonlinear function of a pre-specified vector of option characteristics or "explanatory" variables, $\underbrace{\mathbf{Z}}_{d\times 1} \equiv [F_{t,\tau}, X, \tau, r_{t,\tau}]'$, and uses kernel regression to construct a nonparametric estimate of the function H. The estimator $\hat{H}$ can then be differentiated twice to produce an estimator of the SPD, according to $f_t^*(\cdot) = e^{r_{t,\tau}\tau}\partial^2 H(\cdot)/\partial X^2$. In practice, the dimension of the kernel regression can be reduced by using a semi-parametric approach. The d-dimensional vector of explanatory variables $\mathbf{Z}$ is partitioned into $[\tilde{\mathbf{Z}}', F_{t,\tau}, r_{t,\tau}]'$ where $\tilde{\mathbf{Z}} \equiv [X, F_{t,\tau}, \tau]'$ con-

tains $\tilde{d} = 3$ regressors. Suppose that the call pricing function is given by the parametric Black-Scholes formula except that the implied volatility parameter for that option is a non-parametric function $\sigma(\tilde{\mathbf{Z}})$:

$$\begin{aligned} H(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}) &= H_{BS}(F_{t,\tau}, X, \tau, r_{t,\tau}; \sigma(X, F_{t,\tau}, \tau)) \\ &= e^{-r_{t,\tau}\tau}(F_{t,\tau}\Phi(d_1) - X\Phi(d_2)), \end{aligned}$$

with $d_1 \equiv (\log(F_{t,\tau}/X) + (\sigma^2/2)\tau)/(\sigma\sqrt{\tau})$ and $d_2 \equiv d_1 - \sigma\sqrt{\tau}$. $\sigma \equiv \sigma(X, F_{t,\tau}, \tau)$ is estimated by the Nadaraya-Watson kernel estimator. The implied futures, $F_{t,\tau}$ is derived from using the spot-future parity and the put-call parity relation,

$$H(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}) + Xe^{-r_{t,\tau}\tau} = G(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}) + F_{t,\tau}e^{-r_{t,\tau}\tau},$$

where G denotes the put price at the same strike price X and time-to-expiration τ with the call. Following Aıt-Sahalia & Lo (1998) and Aıt-Sahalia et al. (2001), the future price is inferred from the closest to at-the-money pair of calls and puts. All the prices of in-the-money call options (illiquid) are then replaced using prices of out-of-money puts (liquid) through the put-call parity relationship. Using the index option data of the entire month, I do above procedure on every day t for all available maturities τ to estimate the implied volatility. The state price density is inferred using Black-Scholes formula with τ fixed to be one month.

$Figure.4.a-b$ compares the state-price density with the combined density forecast using all 30 models ($Comb.All$). The blue real-line represents the state-price density estimated by the semi-parametric approach (Method II). The blue dash-line is the state-price density recovered from the call prices directly (Method I). The red real-line represents the physical density predicted by the density forecast. Due to the difference between physical measure and risk-neutral measure, the state price density exhibits fatter tails than the proposed density

forecast. *Figure*.4.*a* shows the forecast made at the end of November, 2007 for December, 2007, when the US stock market just entered a pronounced decline after the peak in October. *Figure*.4.*b* shows the forecast made at the end of December, 2008 for January, 2009, right after the financial institution crisis hits its peak[11].

3 A Portfolio Study

In general, investors require an estimate of the entire distribution of future returns to make their portfolio decisions. In classical asset pricing models, the return distribution is often assumed to be lognormal and time-invariant. In reality, there is strong evidence that the stock returns are fat-tailed and exhibit left-skewness. As an alternative, a portfolio manager may use the empirical return distribution formed by combining various density forecasts to make the optimal portfolio decision. Such empirical return distributions have no closed functional form, and will vary from period to period as the forecast updates according to new information.

This section explores the impact of combined density forecasts on portfolio choice. The purpose of this study is to find a density forecasting method that not only does a better job of modeling return distribution, but also allows the manager to capture portfolio returns that are unrecognized when using the normal assumption.

3.1 An Asset Pricing Model with Empirical Return Distribution

The portfolio study considers both the no-short-sale case and the short-sale-allowed case. If the short sale is not permitted, an investor allocates a_t fraction of total wealth to stocks and the remainder, $(1 - a_t)$ fraction to a risk-free asset, in which case, $0 < a_t \leq 1$. If the

[11]Several major institutions either failed, were acquired under stress, or were subject to government takeover, including Lehman Brothers, Merrill Lynch, Fannie Mae, Freddie Mac, Washington Mutual, Wachovia, Citigroup, and AIG during October and November in 2008.

short sale is allowed, the investor may borrow the risky asset and sell it immediately when the investor anticipates the price of the risky asset will fall; in which case, $a_t < 0$, and the investor invests $(1 - a_t)$ fraction of wealth in the risk-free asset. Let W_t denote the wealth level in period t. The initial wealth W_0 is set to be 1. The budget constraint is given by

$$\begin{aligned} W_{t+1} &= W_t\big[1 + a_t R_{t+1} + (1 - a_t)R_{f,t}\big] \\ &\equiv W_t\big(1 + a_t R^e_{t+1} + R_{f,t}\big). \end{aligned}$$

R_{t+1} is the $S\&P500$ monthly gross return[12]. $R_{f,t}$ is the risk-free rate, for which I use the 1-month Treasury bill rate. R^e_{t+1} is the excess return. The investor is assumed to have the power utility (CRRA) defined over next period's wealth.

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

where γ is the investor's coefficient of relative risk aversion [13]. Portfolio weights for period t can be obtained as the solution to the following optimization problem:

$$a^*_t = \arg\max_{a_t} \int_{-\infty}^{+\infty} \frac{[W_t(1 + a_t R^e_{t+1} + R_{f,t})]^{1-\gamma}}{1-\gamma} f(R^e_{t+1}|\mathcal{F}_t) dR^e_{t+1}. \tag{15}$$

$f(R^e_{t+1}|\mathcal{F}_t)$ can be estimated by the density combination method as in:

$$f(R^e_{t+1}|\mathcal{F}_t) = |\frac{100}{R_{t+1}+1}| \cdot f(y_{t+1}|\mathbf{x}_t, \mathbf{y}_t, \mathbf{w}^*_t) = \sum_{m=1}^{M} w^*_{t,m} f(y_{t+1}|\mathbf{x}_t, \mathbf{y}_t, A_m) \cdot |\frac{100}{R_{t+1}+1}|.$$

For each possible realization range of the next period's regular stock return, its predictive density is the weighted sum of the densities over its counterpart of the continuously compounded return predicted by various models, scaled by the density transformation term.

[12] R_t is the regular return, which equals $(P_t - P_{t-1})/P_t$. y_t is the continuously compounded return, $y_t = 100 \log(P_t/P_{t-1})$. The two forms of return are connected by $y_t = 100 \cdot \log(R_t + 1)$.

[13] For $\gamma = 1$, logarithmic utility is used: $U(W_{t+1}) = \log W_{t+1}$.

When jointly considering the predictive densities obtained in the same way for all possible realizations, one can obtain an empirical return distribution - and hence a density forecast of the future return in *ex-ante*. In particular, let $\mathbf{r} = [\mathbf{r_0}, \ldots, \mathbf{r_S}]$ be a sequence of possible realizations of R^e_{t+1}. The integration in (15) can be approximated by:

$$a^*_t = \arg\max_{a_t} \sum_{i=1}^{\mathbf{S}} \frac{[W_t(1 + a_t R^e_{t+1} + R_{f,t})]^{1-\gamma}}{1-\gamma} P(\mathbf{r}_{i-1} < R^e_{t+1} \leq \mathbf{r}_i | \mathcal{F}_t).$$

As different distributions lead to different estimations of expected return, different density forecasts lead to different portfolio choice, $\{a^*_t\}^T_{t=q+1}$. In solving the optimization, different constrains are imposed depending on whether the short-sale is permitted or if there is margin restriction.

If the short-sale is forbidden, a^*_t must be on the unit interval. The investor can neither borrow shares nor cash but invest using only her own wealth. If the short-sale is permitted, a^*_t is set between $-1 \leq a^*_t \leq 1$. $a^*_t < 0$ means that the investor short sells shares at t to invest in risk-free asset. $a^*_t > -1$ implies that the value of shares borrowed at t cannot exceed the the investor's wealth. To impose a maintenance margin of $\delta\%$ on the short-sale, the equity in the investor's account must be at least $\delta\%$ of the value of her short-position. The account portfolio choice in each period must satisfy

$$\frac{(1-a_t)W_t(1+R_{f,t}) - (\frac{-a_t W_t}{P_t}) \cdot E_t(P_{t+1})}{(\frac{-a_t W_t}{P_t}) \cdot E_t(P_{t+1})} \geq \delta\%,$$

in which, the short-sale cash proceeds, $(1-a_t)W_t$, are invested in the risk-free asset and grow by $(1+R_{f,t})$ by the end of $t+1$; the expected value of the shares the investor must pay back at the end of $t+1$ is $(\frac{-a_t W_t}{P_t}) \cdot E_t(P_{t+1})$. The constraint can be simplified to $\big[(R_{f,t} - \delta\%) - (1+\delta)E_t(R_{t+1})\big] \cdot a_t \leq (1 + R_{f,t})$.

The resulting portfolio weights, a_t^*, give rise to a realized utility next period

$$U(W_{t+1}^*) = \frac{[W_t(1 + a_{t+1}^* R_{t+1}^e + R_{f,t})]^{1-\gamma}}{1-\gamma}.$$

The economic value of the density forecasts can be measured by the certainty equivalent rate of return (CER), which is the utility score of the risky portfolio:

$$CER = \big[(1-\gamma)\frac{1}{T-q}\sum_{t=q+1}^{T} U(W_t^*)\big]^{1/(1-\gamma)} - 1.$$

The certainty equivalent rate is the return rate that risk-free investments would need to offer to provide the same utility score as the risky portfolio. In other words, it is the rate that, if earned with certainty, would provide a utility score equivalent to that of the portfolio in question. CER here compares the utility values of competing density forecasts. The higher the CER, the more attractive is the portfolio strategy, and hence the better is the corresponding density forecast for a risk-averse investor. In the empirical findings, combined density forecasts that assimilate information from the macro-finance variables yield the highest certainty equivalent return among all forecasts.

Meanwhile, density forecasts can provide forward-looking measure of portfolio risk[14]. At the simplest level, one may use the mean-variance framework (Sharpe Ratio) to measure risks. However, the standard deviation can be very unsatisfactory risk measure when dealing with highly non-normal distributions. I hence also examined two alternative risk measures of each density forecast, using the **value at risk (VaR)** and the **expected shortfall (ES)**.

Value at risk (VaR) can be interpreted as the cutoff point such that a loss will not happen with probability greater than p, say, $p = 90\% \ldots 99\%$. VaR is defined as the deviation

[14]For a simple portfolio as above, market risk is the main source of risk.

between the expectation and the $(1-p)$th quantile,

$$VaR_{t+1}(p) = E_t[R^e_{t+1}|\mathcal{F}_t] - Q_{1-p}(R^e_{t+1}|\mathcal{F}_t).$$

VaR usually rises when the confidence level increases. The increasing rate is a point that portfolio managers care to note.

The expected shortfall (ES) is the expected value of the worst $(1-p)\%$ of returns, or to say, the p-th highest losses beyond VaR:

$$ES_{t+1}(p) = E_t[-R^e_{t+1}|R^e_{t+1} < -VaR_{t+1}(p); \mathcal{F}_t].$$

Specifically,

$$\begin{aligned} ES_{t+1}(p) &= \frac{1}{1-p}\int_{-\infty}^{Q_{1-p}(R^e_{t+1}|\mathcal{F}_t)} R^e_{t+1} f[R^e_{t+1}|\mathcal{F}_t] dR^e_{t+1} \\ &= \frac{1}{1-p}\sum_{p=90\%}^{99\%} [\text{p-th highest loss}] \times [\text{probability of p-th highest loss}]. \end{aligned}$$

The ES tells us what to expect in bad states. It gives an idea of how bad the bad might be, while VaR tells us nothing other than to expect a loss higher than VaR itself.

Besides, density forecasts may perform differently in forecasting the different regions of the return distribution. Although portfolio choice is made based on the expected return, a high-quality density forecast should model the tails precisely to allow investors to take risk smartly as well as to prevent loss in *ex-ante*. If a density forecast is more precise in predicting the right tail, it can capture the market surge and hence encourage the investor to long more of the risky asset to obtain the gain. If a density forecast shows more precision on the left tail, it can capture the downward movement of the market, and hence remind the investor to short sell the risky asset to avoid loss. A brief look at the Table 2.*c* shows that combined forecast using macro-finance variables significantly outperforms other forecasts on

the **upper-right tail**. Combined forecast using all 30 models significantly outperforms other forecasts on the **lower-left tail**.

3.2 Empirical Findings

The data and the forecasting specification follow that in Section 2.3 and 3.1. *Table*.3.*a* compares the CERs obtained by portfolios that constructed using different density forecasts, with respect to the increase of relative risk aversion γ. The reported values are the difference in CER between the forecast of interest and the benchmark forecast, EGARCH Student-t($i = 1, j = 1$). Density forecasts that use macro-finance variables can help investors to obtain a high level of certainty equivalent return (CER), especially when the investors are less risk-averse. Except for that the combined forecast of macro-finance variables alone (*Comb.MF*) is of interest, two other representative forecasts are used as analogues: first, the combined forecast of all 30 models (*Comb.All*) combines multiple data generating processes as well as various sources of information. It reflects the premise of forecast combination and hence serves as one analogue. Second, the combined forecast of realized volatility models (*Comb.RV*) has the best performance among all volatility-based models. It uses return information alone but combines multiple data generating processes. It serves as another analogue. The questions of interest are: do density forecasts that use macro-finance variables bring real economic gain, and why?

First, as investors become more risk averse, the difference in gains brought by different density forecasts narrows, and the portfolio choices made by users of different forecasts converge. As indicated by *Table*.3.*a*, for example, when the risk aversion level is low ($\gamma = 1$) and the short-sale is not allowed, the portfolio that uses the combined forecast of all 30 models yields a CER that is up to 0.35% per month higher than that yielded by portfolio that uses the benchmark forecast. As the risk aversion increases from 2 to 100, the difference decreases from 0.35% to 0.00% per month. When the investors

are extremely risk averse, say $\gamma = 100$ or 200, the CERs of all forecasts converge to 0.37% per month, regardless of whether short-sales are allowed or not. This is also the CER provided by always investing in risk-free asset. Meanwhile, portfolio management strategies can be identified by the fraction of wealth invested in the risky assets. The difference between portfolio choices is defined by taking the difference of $a_t^{*,\text{Any Forecast}}$ with $a_t^{*,Comb.MF}$. $Figure$.3.1 shows that as the investors become more risk averse, the variance of the difference in portfolio choices declines, and the ratio allocated to risky assets converges, regardless of whether short-sales are allowed or not.

Second, combined forecasts that use macro-finance variables can yield a certainty equivalent return that is up to 0.09% **per month higher than can be obtained with the combined forecast of realized-volatility models. The advantage of Macro-finance variables is tested through a few turmoil periods of the stock market.** As reported in $Table$.3.a, when the investor is less risk averse ($\gamma \leq 2$), the combined forecasts that use macro-finance variables, including $Comb.MF$ and $Comb.All$, outperform the combined forecast of models that use return information alone, including the combined forecast of realized volatility models ($Comb.RV$), the combined forecast of stochastic volatility models ($Comb.SV$) and the combined forecasts of EGARCH models ($Comb.EGARCH$). The blue shading regions in $Figure$.3.2 show that density forecasts that assimilate macro-finance variables outperform $Comb.RV$ mainly during the period after January, 1984, which is usually referred as "the Great Moderation". The advantage of using density forecasts that incorporate macro-finance variables is tested through the oil crisis in 1970s, the market crash in 1987, the LTCM collapse in 1998, and the recent financial crisis in 2008. Meanwhile, as in $Table$.3.a, the combined forecasts of all 30 models ($Comb.All$) yield a certainty equivalent return that is 0.09% per month higher than can be obtained with the combined forecast of realized volatility models ($Comb.RV$) when $\gamma = 1$ and the short sale is permitted. The combined forecast that use macro-finance variables alone($Comb.MF$) can yield a certainty equivalent return that is from 0.04% to 0.08% per month higher than can be obtained with

$Comb.RV$, regardless of whether short-sales are allowed or not. As $Comb.RV$ yields the highest certainty equivalent return among all forecasts that use return information alone, $Table.3.b$ counts the time frequency when other forecasts surpass $Comb.RV$. Given $\gamma \leq 2$, $Comb.All$ outperforms $Comb.RV$ over 60% of the time over the period from January, 1960, to December, 2011; $Comb.MF$ outperforms $Comb.RV$ over 70% of the time. Such advantage is larger when the short-sale is not allowed.

Third, combined forecasts that use macro-finance variables are helpful to portfolio management because they can better capture the market surge. $Table.2.c$ compares the predictive accuracy on a particular region of interest of the three representative combined density forecasts: $Comb.All$, $Comb.RV$ and $Comb.MF$. Combined density forecast of macro-finance variables ($Comb.MF$) significantly outperforms all other forecasts on the right tail ($\tau \geq 90\%$) of the distribution. Combined density forecast of all 30 models ($Comb.All$) significantly outperforms most forecasts on the left tail ($\tau \leq 10\%$) of the distribution. Combined density forecast of realized volatility models ($Comb.RV$) significantly outperforms most forecasts on the shoulder ($20\% \geq \tau \geq 80\%$) of the distribution. Combined forecast of all 30 models ($Comb.All$) assimilates the advantage of both $Comb.MF$ and $Comb.RV$. It performs the second-best to $Comb.MF$ on the right tail and the second-best to $Comb.RV$ on the shoulder, with a slightly better performance than $Comb.RV$ on the left tail. Using forecasts that assimilate information extracted from macro-finance variables encourages investors to invest more wealth in risky asset whenever there is predictable market surge.

In addition, a simple comparison of the risk measures lead by different density forecasts reveals more facts on the left tail (See $Figure.3.3.a-b$ and $Figure.3.4.a-b$). First, I compare the value-at-risk (VaR) formed by combined forecast of realized volatility models ($Comb.RV$) and combined forecast of all 30 models ($Comb.All$) to that formed by the combined forecast using macro-finance variables ($Comb.MF$). I consider ten confidence levels ($90\% - 99\%$) and test the hypothesis: $VaR_{\text{Any Forecasts},t} - VaR_{Comb.MF,t} = 0$. A negative significant test

statistic at lower confidence levels ($< 97\%$) and a positive significant test statistic at higher confidence levels ($> 98\%$) together indicate that the predictive distributions constructed using volatility-based density forecasts have longer left tails than those constructed using $Comb.MF$. Second, I compare the expected shortfall yield by $Comb.All$ and $Comb.RV$ to $Comb.MF$ in real time. The positive significant t-test statistics over ten confidence levels ($90\% - 99\%$) show that the Expected Shortfall estimated by forecasts that assimilate the market volatility is significantly larger than forecasts that assimilate macro-finance variables alone. In this sense, volatility-based models contribute to model the left-tail, whereas macro-finance models contribute to model the right tail[15]. A combination of them captures the left-skewness as well as the potential market surge.

4 Conclusion

This paper finds that the combined density forecast that uses various sources of information is significantly more accurate than the combined density forecast that uses a single source of information, in forecasting the U.S. $S\&P$ 500 index monthly returns from 1950 to 2011 in real-time. This paper contributes to the empirical work on density forecast combination, especially that of Durham & Geweke (2011), by providing empirical evidence that combining quantile density forecasts in way that maximizes the log predictive likelihood exhibits better forecasting performance than SVOL models and EGARCH models. This paper also contributes to the stock return forecasting literature, such as Welch & Goyal (2008), by finding that macro-finance variables exhibit substantial predictive power in forecasting the stock return distribution. Furthermore, the portfolio study shows that a better density forecast that incorporates information carried by macro-finance variables encourages the investor to take moderate risk and can obtain substantial economic gains in real-time.

[15]Unless when extremely risk averse, portfolio managers who use density forecasts with macro-finance variables alone may invest significantly heavily in the risky asset. $Figure.3.3.c$ and $Figure.3.4.c$ use a simple t-test to compare whether $a^*_{Comb.MF,t}$ is significantly larger than $a^*_{Comb.All,t}$ and $a^*_{Comb.RV,t}$.

A Estimation of Stochastic Volatility Model

This section illustrates the estimation procedure of the stochastic volatility models. Following Kim et al. (1998) and Omori et al. (2007), the conditional mean equation was first transformed into a linear model by taking logarithm of the squares of observations:

$$\begin{aligned} y_t^* &\equiv \log y_t^2 - \log(\lambda_t) = h_t + \log \varepsilon_t^2 \\ d_t &= sign(y_t) = I(\varepsilon_t > 0) - I(\varepsilon_t \leq 0). \end{aligned}$$

Let $\xi_t = \log \varepsilon_t^2$. The first equation can be rewritten as

$$y_t^* = h_t + \xi_t.$$

In the basic SVOL model, ξ_t follows a $\log \chi^2$ distribution. Kim et al. (1998) approximate this distribution by a mixture of seven Gaussian distributions to match its first four moments. In particular,

$$g(\xi_t) = \sum_{i=1}^{K} p_i f_N(\xi_t | m_i, v_i^2), \tag{16}$$

where $K = 7$ and $f_N(\xi_t | m_i, v_i^2)$ denotes the density function of a normal distribution with mean m_i and variance v_i^2.

In the SVOL model with leverage effect, Omori et al. (2007) approximate the bivariate conditional density of $(\xi_t, \eta_t | d_t)$ by a ten components mixture of bivariate Gaussian densities.

$$g(\xi_t, \eta_t) = \sum_{i=1}^{K} p_i f_N(\xi_t | m_i, v_i^2) f_N[d_t \rho \sigma \exp(m_i/2) a_i + b_i(\xi_t - m_i), \sigma^2(1-\rho^2)]. \tag{17}$$

The constants m_i and v_i^2 are determined on the basis of $K = 10$ components. The selection of $(p_i, m_i, v_i^2, a_i, b_i)$ is listed in the following table.

	Selection of $(p_i,\ m_i,\ v_i^2,\ a_i,\ b_i)$				
N = 10	p_i	m_i	v_i^2	a_i	b_i
$i = 1$	0.00609	1.92677	0.11265	1.01418	0.50710
$i = 2$	0.04775	1.34744	0.17788	1.02248	0.51124
$i = 3$	0.13057	0.73504	0.26768	1.03403	0.51701
$i = 4$	0.20674	0.02266	0.40611	1.05207	0.52604
$i = 5$	0.22715	−0.85173	0.62699	1.08153	0.54076
$i = 6$	0.18842	−1.97278	0.98583	1.13114	0.56557
$i = 7$	0.12047	−3.46788	1.57469	1.21754	0.60877
$i = 8$	0.05591	−5.55246	2.54498	1.37454	0.68728
$i = 9$	0.01575	−8.68384	4.16591	1.68327	0.84163
$i = 10$	0.00115	−14.65000	7.33342	2.50097	1.25049

Through above steps, the stochastic volatility model is approximated by a linear Gaussian state space model conditioned on the mixture component indicators $s_t \in \{1, 2, \ldots, K\}$. It then becomes possible to efficiently sample the posterior distribution of $\mathbf{s} = \{s_t\}_{t=1}^T$, the volatility component $\mathbf{h} = \{h_t\}_{t=1}^T$ and the parameters $\theta = \{\mu, \phi, \sigma, \rho\}$ by MCMC methods.

1. **Sampling volatility parameters** $(\theta, \mu, \mathbf{h})$

- The conditional posterior probability density function of $(\theta, \mu, \mathbf{h})$ is

$$\pi(\theta, \mu, \mathbf{h}|\mathbf{s}, \mathbf{y}^*, \mathbf{d}) \propto \pi(\theta|\mathbf{s}, \mathbf{y}^*, \mathbf{d})\pi(\mu, \mathbf{h}|\mathbf{s}, \mathbf{y}^*, \mathbf{d})$$

where

$$\begin{aligned} \pi(\theta|\mathbf{s}, \mathbf{y}^*, \mathbf{d}) &\propto f(\mathbf{y}^*|\theta, \mathbf{s}, \mathbf{d})\pi(\theta) \\ \pi(\mu, \mathbf{h}|\mathbf{s}, \mathbf{y}^*, \mathbf{d}) &\propto \pi(\mu|\theta, \mathbf{s}, \mathbf{y}^*, \mathbf{d})\pi(\mathbf{h}|\mu, \theta, \mathbf{s}, \mathbf{y}^*, \mathbf{d}) \end{aligned}$$

The conditional likelihood $f(\mathbf{y}^*|\theta, \mathbf{s}, \mathbf{d})$ is evaluated through the augmented Kalman filter (see Appendix B). $\pi(\theta)$ uses the prior distribution:

$$\frac{\phi+1}{2} \sim Beta(20, 1.5), \quad \sigma^2 \sim \mathcal{IG}(5/2, 0.05/2), \quad \frac{\rho+1}{2} \sim Beta(1,1).$$

The next step is to find $\hat{\theta} = (\hat{\phi}, \hat{\sigma}, \hat{\rho})$ that maximizes (or approximately maximizes) the posterior probability density $\pi(\theta|\mathbf{s}, y^*, \mathbf{d})$ and generate a candidate θ^* from a normal distribution $N(\theta_*, \Sigma_*)$, truncated over $R = \{\theta : |\phi| < 1, \sigma > 0, |\rho| < 1\}$, where

$$\theta_* = \hat{\theta} + \Sigma_* \frac{\partial \log \pi(\theta|\mathbf{s}, \mathbf{y}^*, \mathbf{d})}{\partial \theta}\Big|_{\theta=\hat{\theta}}, \quad \Sigma_*^{-1} = -\frac{\partial \log \pi(\theta|\mathbf{s}, \mathbf{y}^*, \mathbf{d})}{\partial \theta \partial \theta'}\Big|_{\theta=\hat{\theta}}.$$

Let θ_0 denote the last draw of θ. We accept the candidate θ^* with probability.

$$a(\theta_0, \theta^*|\mathbf{s}, \mathbf{y}^*, \mathbf{d}) = \min\Big\{\frac{\pi(\theta^*|\mathbf{s}, \mathbf{y}^*, \mathbf{d}) f_N(\theta_0|\theta_*, \Sigma_*)}{\pi(\theta_0|\mathbf{s}, \mathbf{y}^*, \mathbf{d}) f_N(\theta^*|\theta_*, \Sigma_*)}, 1\Big\},$$

where f_N denotes the density of the truncated normal distribution for the proposal above. If the candidate θ^* is rejected, the current value θ_0 is taken as the next draw.

- To sample $(\mu, \mathbf{h})|\theta, \mathbf{s}, \mathbf{y}^*, \mathbf{d}$, I first generate $\mu|\theta, \mathbf{s}, \mathbf{y}^*, \mathbf{d} \sim N(Q_{n+1}^{-1} q_{n+1}, Q_{n+1}^{-1})$, where q_{n+1} and Q_{n+1} are the byproducts of the augmented Kalman Filter (see Appendix B). I then sample $\mathbf{h}|\mu, \theta, \mathbf{s}, \mathbf{y}^*, \mathbf{d}$ in one block using the simulation smoother. Given $s_t = i$, the approximating linear Gaussian state space model is formed by

$$\begin{aligned} y_t^* &= m_i + h_t + G_t u_t, \\ h_{t+1} &= d_t \rho \sigma a_i \exp(m_i/2) + (1-\phi)\mu + \phi h_t + H_t u_t, \end{aligned}$$

where $u_t \sim N(0, I_2)$, $G_t = (v_i, 0)$, and $H_t = (d_t \rho \sigma b_i v_i \exp(m_i/2), \sigma\sqrt{1-\rho^2})$. The prior of μ uses $\mu \sim N(-10, 1)$.

2. Sampling heavy-tailed parameters $(\lambda, \nu)|\theta, \mu, \mathbf{s}, \mathbf{h}, \mathbf{y}$

- In the basic SVOL model, λ_t is set to be 1 for all time periods and the algorithm directly enters sampling the mixture state without sampling ν. In the fat-tailed SVOL model, the heavy-tailed parameters are sampled from the conditional posterior distribution of (λ, ν), where the joint probability density is

$$\begin{aligned}\pi(\lambda, \nu|\theta, \mu, \mathbf{s}, \mathbf{h}, \mathbf{y}) &\propto f(\mathbf{y}|\theta, \mu, \mathbf{s}, \mathbf{h}) g(\lambda|\nu)\pi(\nu) \\ &\propto \pi(\nu)\Pi_{t=1}^{T} \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} \lambda_t^{-(\frac{\nu}{2}+1)} \exp\big\{ -\frac{\nu}{2\lambda_t} - \frac{(\log\lambda_t - \mu_{\lambda_t})^2}{2\sigma_{\lambda_t}^2}\big\},\end{aligned}$$

where

$$\begin{aligned}\mu_{\lambda_t} &= \log y_t^2 - m_i - h_t - \frac{d_t\rho b_i v_i \exp(m_i/2)\{h_{t+1} - \phi h_t - (1-\phi)\mu - d_t\rho\sigma a_i \exp(m_i/2)\}}{\sigma\{(1-\rho^2) + \rho^2 b_i^2 v_i^2 \exp(m_i)\}}, \\ \sigma_{\lambda_t}^2 &= \frac{v_i^2(1-\rho^2)}{1-\rho^2+\rho^2 b_i^2 v_i^2 \exp(m_i)},\end{aligned}$$

given $s_t = i$, for $t = 1, \ldots, T-1$ and $\mu_{\lambda_T} = \log y_T^2 - m_i - h_T$, $\sigma_{\lambda_T}^2 = v_i^2$. The conditional posterior distribution for λ_t is given by

$$\pi(\lambda|\theta, \mu, \nu, \mathbf{s}, \mathbf{h}, \mathbf{y}) \propto \lambda_t^{-(\frac{\nu}{2}+1)} \exp\big\{ -\frac{\nu}{2\lambda_t} - \frac{(log\lambda_t - \mu_{\lambda_t})^2}{2\sigma_{\lambda_t}^2}\big\},$$

Then λ_t is sampled using the M-H algorithm with the candidate drawn by $(\lambda_t)^{-1} \sim Gamma(\nu/2, \nu/2)$.

- The conditional posterior distribution for ν is given by

$$\pi(\nu|\lambda) \propto \pi(\nu)\frac{\big(\frac{\nu}{2}\big)^{\frac{T\nu}{2}}}{\Gamma(\frac{\nu}{2})^T}\Pi_{t=1}^{T}\lambda_t^{-\frac{\nu}{2}} \exp\big(-\frac{\nu}{2}\sum_{t=1}^{T}\lambda_t^{-1}\big).$$

The conditional prior of $\pi(\nu)$ uses $\nu \sim Gamma(16, 0.8)$. Let ν denote the mode (or approximate mode) of the conditional posterior density $\pi(\nu|\lambda)$, and let $l(\nu) =$

$\log \pi(\nu|\lambda)$. Applying Taylor expansion to $l(\nu)$ around $\hat{\nu}$ as

$$l(\nu) \approx l(\hat{\nu}) + l'(\hat{\nu})(\nu - \hat{\nu}) + \frac{1}{2} l''(\hat{\nu})(\nu - \hat{\nu})^2 \equiv h(\nu),$$

where $l'(\nu)$ and $l''(\nu)$ are the first and second derivative of $l(\nu)$ evaluated at $\nu = \hat{\nu}$. The approximating density $N(\mu_\nu, \sigma_\nu^2)$is truncated over $(0, \infty)$, where $\mu_\nu = \hat{\nu} - l'(\nu)/l''(\nu)$ and $\sigma_\nu^2 = -1/l''(\nu)$. Finally, ν is sampled by the following two steps.

i. A-R step: Generate a candidate $\nu^* \sim N(\mu_\nu, \sigma_\nu^2)$ truncated over $(0, \infty)$ and accept ν^* with probability $\min(1, \exp\{l(\nu^*) - h(\nu^*)\})$. If it is rejected, generate ν^* again until the candidate is accepted.

ii M-H step: Let ν_0 denote the current point of ν. Accept ν^* with probability

$$\min\Big\{\frac{\exp[l(\nu^*)]\min\{\exp[(l(\nu_0)], \exp[h(\nu_0)]\}}{\exp[l(\nu_0)]\min\{\exp[l(\nu^*)], \exp[h(\nu^*)]\}}, 1\Big\}.$$

If ν^* is rejected, ν_0 is retained as the new draw.

3. Sampling mixture state s To sample s_t, one simply computes

$$\pi(s_t = i|\theta, \mu, \mathbf{h}, \mathbf{y}^*, \mathbf{d}) \propto q_i \frac{1}{v_i} \exp\Big\{-\frac{(y_t^* - h_t - m_i)^2}{2v_i^2}\Big\} \exp\Big[-\frac{\{(h_{t+1} - \mu) - \phi(h_t - \mu) - D_i(y_t^*)\}^2}{2\sigma^2(1-\rho^2)}\Big],$$

for $i = 1, \ldots, K$, in which $y_t^* = \log y_t^2 - \log \lambda_t$ and

$$D_i(y_t^*) = d_t \rho \sigma \{a_i + b_i(y_t^* - h_t - m_i)\} \exp(\frac{m_i}{2}).$$

s_t is sampled from the K-point discrete distribution independently for $t = 1, \ldots, T$. In the case of $t = T$, the second $\exp[\cdot]$ is omitted. In the MCMC sampling of the posterior distribution, the initial 1000 sweeps are discarded and the subsequent 1500 sweeps are retained for analysis.

B Augmented Kalman filter

To find θ that maximizes $\pi(\theta|\mathbf{s}, \mathbf{y}^*, \mathbf{d})$, one needs to evaluate a likelihood $f(\mathbf{y}^*|\theta, \mathbf{s}, \mathbf{d})$. Assuming $\theta = (\phi, \sigma, \rho)$ is fixed, one may conduct an augmented Kalman filter and calculate the likelihood function.

The approximating state space model after the log-quadratic transformation of the SVOL model is,

$$\begin{aligned} y_t^* &= m_i + h_t + G_t u_t, \quad t = 1, \ldots, T, \\ h_{t+1} &= d_t \rho \sigma a_i \exp(m_i/2) + \mu(1-\phi) + \phi h_t + H_t u_t, \quad t = 1, \ldots, T, \end{aligned}$$

where $\mu \sim N(-10, 1)$, and $u_t \sim N(0, 1)$. Let $h_{t+1|t}$ and $P_{t+1|t}$ denote the predictive mean and variance of h_t. $h_{1|0} = 0$ and $A_{1|0}^* = 1$. When μ is fixed, the Kalman filter is the recursion.

$$\begin{aligned} h_{t+1|t} &= b_t + \phi h_{t|t-1} + f_t K_t, \\ P_{t+1|t} &= \phi P_{t|t-1} \phi' - \phi P_{t|t-1} K_t' + H_t (H_t - K_t G_t)'. \end{aligned}$$

H_t and G_t are defined at the end of sampling volatility parameters, and

$$\begin{aligned} b_t &= d\rho\sigma a_{it} \exp(m_{it}/2), \qquad f_t = y_t^* - m_{it} - h_{t|t-1} \\ K_t &= (\phi P_{t|t-1} + H_t G_t') D_t^{-1}, \qquad D_t = P_{t|t-1} + G_t G_t'. \end{aligned}$$

The log-likelihood of $\mathbf{y}$ given μ is

$$\log f(\mathbf{y}|u) = -\frac{1}{2}\big\{T \log 2\pi + \log|\Sigma| + (\mathbf{y} - m_i)'\Sigma^{-1}(\mathbf{y} - m_i) - 2q'\mu + \mu' Q \mu\big\},$$

where $\log|\Sigma| = \sum_{t=1}^T \log|D_t|$, $(\mathbf{y} - m_i)'\Sigma^{-1}(\mathbf{y} - m_i) = \sum_{t=1}^T f_t' D_t^{-1} f_t$, $q = \sum_{t=1}^T F_t' D_t^{-1} f_t$ and

$Q = \sum_{t=1}^{T} F_t' D_t^{-1} F_t$. The posterior distribution of μ given y is $N(Q_{T+1}^{-1} q_{T+1}, Q_{T+1}^{-1})$ where

$$\begin{aligned} q_{t+1} &= q_t + F_t' D_t^{-1} f_t, \\ Q_{t+1} &= Q_t + F_t' D_t^{-1} F_t, \end{aligned}$$

in which $F_t = -A_{t|t-1}$, and $A_{t+1|t} = -(1-\phi) + \phi A_{t|t-1} + K_t F_t$. Thus the likelihood of $\mathbf{y}$ is

$$\begin{aligned} \log f(\mathbf{y}) &= \log f(\mathbf{y}|\mu) + \log \pi(\mu) - \log \pi(\mu|\mathbf{y}) \\ &= const. - \frac{1}{2}\Big\{ \sum_{t=1}^{T} \log |D_t| + \log |Q_{T+1}| + \sum_{t=1}^{T} f_t' D_t^{-1} f_t - q_{T+1}' Q_{T+1}^{-1} q_{T+1} \Big\}. \end{aligned}$$

C ATIC Quantile-varying Factor Selection Rule

As factor selection is found to improve forecasting the mean of a time series (Bai & Ng (2008)), Ando & Tsay (2011) proposed a factor selection method for quantile regression. Each quantile of y_t is predicted by

$$\hat{Q}_\tau(y_t|\mathbf{x}_{t-1}) = \hat{\beta}_0(\tau) + \hat{\beta}_1(\tau) y_{t-1} + \hat{\beta}_2(\tau)\mathbf{f}_{t-1} \equiv \hat{\beta}(\tau)\hat{\mathbf{z}}_{t-1}, \qquad t = 1, \ldots, T.$$

$\mathbf{f}_{t-1}$ is the first $\mathbf{r}$ principal components of $\underbrace{\mathbf{x}_{t-1}}_{N\times 1}$. The *optimal* number of factors $\mathbf{r}^*$ for each quantile is found by minimizing the following information criteria:

$$ATIC = -2\{\eta_\tau(\hat{G}; \hat{\beta}, \hat{\mathbf{Z}}) - \hat{b}_\tau(G)\}, \qquad \tau = 0.01, \ldots, 0.99.$$

$\eta_\tau(\hat{G};\hat{\beta},\hat{\mathbf{Z}})$ is the sample-based log-likelihood of $\mathbf{y}$. $\hat{\mathbf{Z}}$ is the matrix of $\hat{\mathbf{z}}_{t-1}$ for $t = 1,\ldots,T$. $\hat{b}_\tau(G)$ is an estimator of the bias between the sample log likelihood (smaller) and the true expected log-likelihood (larger), which is generally positive. $\eta_\tau(\hat{G};\hat{\beta},\hat{\mathbf{Z}})$ takes the form:

$$\eta_\tau(\hat{G};\hat{\beta},\hat{\mathbf{Z}}) = \int l_\tau(y;\hat{\beta},\hat{\mathbf{Z}})d\hat{G}(y) = \log\left[\tau(1-\tau)\right] - \frac{1}{T}\sum_{t=1}^{T}\rho_\tau(y_t - \hat{\beta}(\tau)\hat{\mathbf{z}}_t),$$

in which $l_\tau(y;\hat{\beta},\hat{\mathbf{Z}})$ to be the log-likelihood function of y. $\hat{\beta}$ is estimated by solving $\partial\eta_\tau(y;\hat{\beta},\hat{\mathbf{Z}})/\partial\beta = 0$.

The bias term $\hat{b}_\tau(G)$ is approximately given by

$$\hat{b}_\tau(G) = \frac{1}{T}tr[J_\tau^{-1}(\hat{\mathbf{Z}})\cdot I_\tau(\hat{\mathbf{Z}})] + \frac{1}{TN}\sum_{t=1}^{T} g(\hat{\xi}_t)tr[K_\tau(\hat{\beta}(\tau))\cdot\hat{\Sigma}_z(t)] + O\big(\frac{1}{B_{N,T}}\big),$$

in which $B_{N,T} = \min\{N, T\sqrt{T}\}$, and

$$\begin{aligned}
I_\tau(\hat{\mathbf{Z}}) &= \frac{1}{T}\tau(1-\tau)\hat{\mathbf{Z}}'\hat{\mathbf{Z}} \\
J_\tau(\hat{\mathbf{Z}}) &= \frac{1}{T}\hat{\mathbf{Z}}'\hat{M}\hat{\mathbf{Z}}. \\
K_\tau(\hat{\beta}(\tau)) &= \hat{\beta}(\tau)\hat{\beta}(\tau)', \\
\hat{\Sigma}_z(t) &= \begin{bmatrix} \tilde{V}^{-1}\hat{Q}_t\tilde{V}^{-1} & \hat{L}_{f,w} \\ \hat{L}'_{f,w} & \hat{\Sigma}_w \end{bmatrix}.
\end{aligned}$$

w refers to the constant and lag values of y_t. $\hat{L}'_{f,w}$ is the covariance of the factors and w. $M = diag\{g(\hat{\xi}_1(\tau)),\ldots,g(\hat{\xi}_T(\tau))\}$ is a T-dimensional diagonal matrix, in which quantile estimates, $\hat{\xi}_t(\tau) = G^{-1}(\tau|\hat{\mathbf{z}}_t)$, and $g(\hat{\xi}_t(\tau))$ is the corresponding normal density. Assuming that ε_{it} is cross-sectionally uncorrelated, $\hat{Q}_t = \frac{1}{N}\sum_{i=1}^{N}\hat{\varepsilon}_{it}^2\hat{\lambda}_i\hat{\lambda}_j{}'$, and $\hat{\varepsilon}_{it} = x_{it} - \hat{\lambda}_i'\hat{\mathbf{f}}_t$. $\hat{\lambda}_i$ is the factor loading of variable $x_{i,t-1}$.

In sum, $\eta_\tau(\hat{G};\hat{\beta},\hat{\mathbf{Z}})$ measures the goodness of fit of the model and $\hat{b}_\tau(G)$ is a penalty that measures the complexity of the model as more factors are included.

Table.1. Predictive Likelihood of Density Forecasts in Five Classes

Sample Periods $1959:01-2011:12$

Class	Combined Forecasts	Individual Forecasts			
I	**Combined**	*Dividend*	*Earnings*	*SVAR*	*Book to Mkt Ratio*
	Macro-finance	51.99	54.56	72.73	67.80
	Variables	*Net Equity Exp.*	*Term Spread*	*Default Yield Spread*	*Inflation*
	76.51	67.60	67.96	75.90	66.61
		Unemployment Rate	*Industrial Production Growth*	*Non-farm Payroll*	
		70.50	77.93	78.13	
II	**Single-model**	*Ando-Tsay Factor*	*1st. P.C.*	*Multivariate*	
	Multivariate Forecasts	66.20	58.89	-7.65	
III	**Combined**	*Gaussian* $(i=1,j=1)$	*Gaussian* $(i=1,j=2)$	*Gaussian* $(i=2,j=1)$	*Gaussian* $(i=2,j=2)$
	EGARCHs	39.48	31.53	22.44	37.31
	53.28	*Student-t* $(i=1,j=1)$	*Student-t* $(i=2,j=1)$	*GED* $(i=1,j=1)$	*GED* $(i=2,j=1)$
		0.00	2.86	15.74	-13.75
IV	**Combined SVs**	*Gaussian SV*	*Fat-tail SV*	*Corr SV*	*Fat-tail Corr SV*
	61.75	59.38	60.00	64.99	55.82
V	**Combined RVs**	*RV-Gaussian* $(j=1)$	*RV-Gaussian* $(j=2)$	*RV-t*	*RV-GED*
	89.99	33.09	32.71	55.68	20.39
	Combine all 30 models				
	90.62				

Note: The benchmark forecast is EGARCH Student-t$(i=1,j=1)$. A difference of 6.4080 generally corresponds to a 1% increment in probability. Using the nonparametric kernel density forecast that use historical returns alone (-1850) yields similar result.

Table.2.a Amisano & Giacomini (2007) Test Results: Combined Density Forecasts

Models	**Comb.RV**	**Comb.MF**	**ATIC**	**Comb.SV**	**1st. PC**	**Comb.EGARCH**	**Multivariate**
Comb.All	0.11	2.07	2.67	2.84	3.71	4.07	5.48
Asy. p-value	(0.46)	(0.02*)	(0.00*)	(0.00*)	(0.00*)	(0.00*)	(0.00*)
Boot. p-value	[0.46]	[0.02*]	[0.01*]	[0.01*]	[0.00*]	[0.00*]	[0.00*]
Comb.RV		1.37	2.23	2.58	3.02	3.22	4.75
Asy. p-value		(0.09)	(0.01*)	(0.00*)	(0.00*)	(0.00*)	(0.00*)
Boot. p-value		[0.08]	[0.02*]	[0.01*]	[0.00*]	[0.00*]	[0.00*]
Comb.MF			1.41	1.27	2.10	2.14	4.73
Asy. p-value			(0.08)	(0.10)	(0.02*)	(0.02*)	(0.00*)
Boot. p-value			[0.08]	[0.11]	[0.02*]	[0.02*]	[0.00*]
ATIC. Factor				0.41	1.15	1.13	3.78
Asy. p-value				(0.34)	(0.12)	(0.13)	(0.00*)
Boot. p-value				[0.34]	[0.11]	[0.12]	[0.00*]
Comb.SV					0.26	0.70	3.33
Asy. p-value					(0.40)	(0.24)	(0.00*)
Boot. p-value					[0.39]	[0.22]	[0.00*]
1st. PC						0.56	3.35
Asy. p-value						(0.29)	(0.00*)
Boot. p-value						[0.27]	[0.00*]
Comb.EGARCH							2.86
Asy. p-value							(0.00*)
Boot. p-value							[0.00*]

Table.2.b Amisano & Giacomini (2007) Test Results: Combine All v.s. Individual

Sample Periods $1959:01-2011:12$

Combined Density Forecasts	**Individual Density Forecasts**			
RVs	*RV-Gaussian* $(j=1)$	*RV-Gaussian* $(j=2)$	*RV-t*	*RV-GED*
AG stat	1.66	1.60	3.69	2.22
Asy. p-value	(0.05^*)	(0.06)	(0.00^*)	(0.01^*)
Boot. p-value	$[0.01^*]$	$[0.03^*]$	$[0.00^*]$	$[0.00^*]$
SVs	*Gaussian SV*	*Fat-tail SV*	*Corr SV*	*Fat-tail Corr SV*
AG stat	2.88	2.90	2.51	3.06
Asy. p-value	(0.00^*)	(0.00^*)	(0.01^*)	(0.00^*)
Boot. p-value	$[0.02^*]$	$[0.03*]$	$[0.00^*]$	$[0.00^*]$
EGARCHs	*Gaussian* $(i=1,j=1)$	*Gaussian* $(i=1,j=2)$	*Gaussian* $(i=2,j=1)$	*Gaussian* $(i=2,j=2)$
AG stat	3.92	2.82	3.59	4.02
Asy. p-value	(0.00^*)	(0.00^*)	(0.00^*)	(0.00^*)
Boot. p-value	$[0.00^*]$	$[0.00^*]$	$[0.00^*]$	$[0.00^*]$
	Student-t $(i=1,j=1)$	*Student-t* $(i=2,j=1)$	*GED* $(i=1,j=1)$	*GED* $(i=2,j=1)$
AG stat	6.59	6.44	3.78	3.49
Asy. p-value	(0.00^*)	(0.00^*)	(0.00^*)	(0.00^*)
Boot. p-value	$[0.00^*]$	$[0.00^*]$	$[0.00^*]$	$[0.00^*]$
Combined MFs	*Dividend*	*Earnings*	*SVAR*	*Book to Mkt Ratio*
AG stat	3.43	3.57	1.85	2.26
Asy. p-value	(0.00^*)	(0.00^*)	(0.03^*)	(0.01^*)
Boot. p-value	$[0.00^*]$	$[0.00^*]$	$[0.03^*]$	$[0.01^*]$
	Net Equity Exp.	*Term Spread*	*Default Yield Spread*	*Inflation*
AG stat	2.34	2.27	1.87	2.14
Asy. p-value	(0.01^*)	(0.01^*)	(0.03^*)	(0.02^*)
Boot. p-value	$[0.01^*]$	$[0.01^*]$	$[0.04^*]$	$[0.00^*]$
	Unemployment Rate	*Industrial Production Growth*	*Non-farm Payroll*	
AG stat	1.94	1.32	1.33	
Asy. p-value	(0.03^*)	(0.09)	(0.09)	
Boot. p-value	$[0.02^*]$	$[0.09]$	$[0.08]$	

Note: $(\cdot)$ is the asymptotic p-values. $[\cdot]$ is the bootstrap p-values. $*$ indicates significance at 5% level.

Table 2.c Diks et al. (2011) Test Results

Sample Periods 1959 : 01 − 2011 : 12

	Lower 10% (Left Tail)					**Upper 10% (Right Tail)**				
Models	**Comb.RV**	**Comb.SV**	**Comb.EGARCH**	**Comb.MF**	**ATIC**	**Comb.RV**	**Comb.SV**	**Comb.EGARCH**	**Comb.MF**	**ATIC**
Comb.All	0.87	2.94	4.33	2.20	2.90	2.13	1.49	1.47	−1.62	0.83
Asy. p-value	(0.19)	(0.00*)	(0.00*)	(0.01*)	(0.00*)	(0.02*)	(0.07)	(0.07)	(0.95)	(0.20)
Models	**Comb.All**	**Comb.SV**	**Comb.EGARCH**	**Comb.MF**	**ATIC**	**Comb.All**	**Comb.SV**	**Comb.EGARCH**	**Comb.MF**	**ATIC**
Comb.RV	−0.87	2.50	3.34	1.26	2.29	−2.13	0.43	0.74	−1.92	0.07
Asy. p-value	(0.19)	(0.01*)	(0.00*)	(0.10*)	(0.01*)	(0.98)	(0.34)	(0.23)	(0.97)	(0.47)
Models	**Comb.All**	**Comb.RV**	**Comb.SV**	**Comb.EGARCH**	**ATIC**	**Comb.All**	**Comb.RV**	**Comb.SV**	**Comb.EGARCH**	**ATIC**
Comb.MF	−2.20	−1.26	1.27	1.91	1.89	1.62	1.92	3.47	1.87	1.81
Asy. p-value	(0.99)	(0.90)	(0.10)	(0.03)	(0.03)	(0.05*)	(0.03*)	(0.00*)	(0.03*)	(0.04*)

	Between 20% and 80% (Shoulder)				
Models	**Comb.RV**	**Comb.SV**	**Comb.EGARCH**	**Comb.MF**	**ATIC**
Comb.All	−0.26	2.69	4.58	2.58	2.77
Asy. p-value	(0.60)	(0.00*)	(0.00*)	(0.00*)	(0.00*)
Models	**Comb.All**	**Comb.SV**	**Comb.EGARCH**	**Comb.MF**	**ATIC**
Comb.RV	0.26	2.72	3.94	1.90	2.49
Asy. p-value	(0.40)	(0.00*)	(0.00*)	(0.03*)	(0.01*)
Models	**Comb.All**	**Comb.RV**	**Comb.SV**	**Comb.EGARCH**	**ATIC**
Comb.MF	−2.58	−1.90	0.83	1.75	0.94
Asy. p-value	(1.00)	(0.97)	(0.20)	(0.04*)	(0.17)

Note: $(\cdot)$ is the asymptotic p-values. $*$ indicates significance at 10% level.

Table.3.a The Difference in CER by December, 2011: % per month

Models	Strategy	Risk Aversion									
		$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$	$\gamma = 5$	$\gamma = 6$	$\gamma = 7$	$\gamma = 10$	$\gamma = 100$	$\gamma = 200$
Comb.All	*no short-sale*	0.16	0.13	0.13	0.11	0.10	0.08	0.07	0.05	0.00	0.00
	short-sale	0.35	0.27	0.27	0.25	0.23	0.20	0.18	0.13	0.00	0.00
	(margin = 0.5)	0.33	0.25	0.25	0.23	0.20	0.17	0.12	0.01	0.00	0.00
Comb.RV	*no short-sale*	0.11	0.12	0.11	0.09	0.08	0.06	0.06	0.04	0.00	0.00
	short-sale	0.26	0.29	0.27	0.25	0.22	0.19	0.17	0.12	0.01	0.00
	(margin = 0.5)	0.25	0.28	0.27	0.25	0.20	0.22	0.29	0.12	0.00	0.00
Comb.MF	*no short-sale*	0.15	0.11	0.07	0.06	0.05	0.05	0.04	0.02	0.00	0.00
	short-sale	0.34	0.27	0.23	0.20	0.18	0.16	0.13	0.09	0.00	0.00
	(margin = 0.5)	0.33	0.26	0.22	0.19	0.17	0.15	0.13	0.09	0.00	0.00
ATIC. Factor	*no short-sale*	0.12	0.12	0.11	0.06	0.05	0.03	0.02	0.02	0.00	0.00
	short-sale	0.20	0.17	0.16	0.14	0.11	0.08	0.06	0.04	0.00	0.00
	(margin = 0.5)	0.21	0.17	0.16	0.15	0.11	0.07	0.06	0.01	0.00	0.00
Comb.SV	*no short-sale*	0.05	0.08	0.09	0.08	0.07	0.06	0.06	0.05	0.00	0.00
	short-sale	0.13	0.14	0.18	0.19	0.17	0.16	0.14	0.11	0.00	0.00
	(margin = 0.5)	0.12	0.14	0.17	0.18	0.15	0.14	0.09	0.02	0.00	0.00
Comb.EGARCH	*no short-sale*	0.10	0.11	0.11	0.09	0.07	0.05	0.04	0.02	0.00	0.00
	short-sale	0.24	0.24	0.23	0.20	0.16	0.14	0.12	0.08	0.01	0.00
	(margin = 0.5)	0.20	0.20	0.21	0.19	0.15	0.14	0.11	0.07	0.01	0.00

Note: $CER = \left[(1-\gamma)\frac{1}{T-q}\sum_{t=q+1}^{T} U(W_t^*)\right]^{1/(1-\gamma)} - 1$, where $T = 744$ and $q = 120$. In no-short sale case, CER of the benchmark model fluctuates between 0.36 and 0.39 and converges to 0.37. In short-sale allowed case, CER of the benchmark model fluctuates between 0.17 and 0.26 and converges to 0.37.

Table.3.b Comparison of the CER: Frequency Count

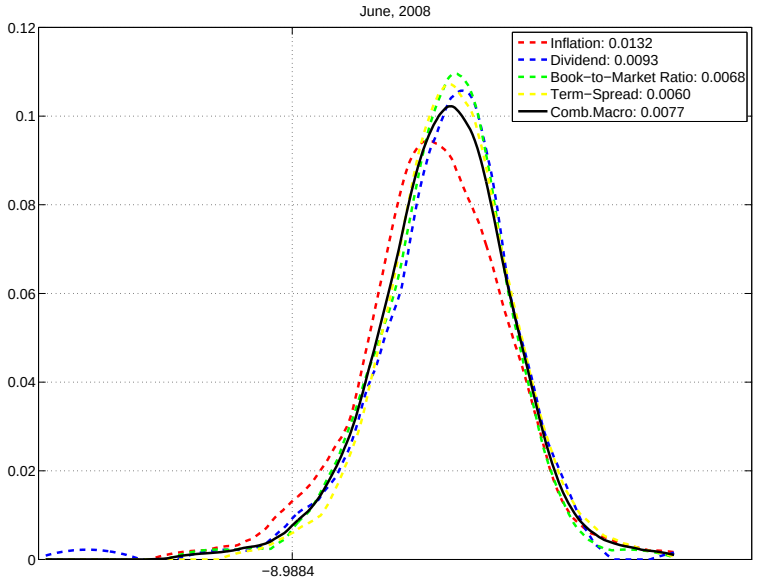
Comb.Any Forecast > Comb.RV, Unit: % of time

Sample Periods: $1959:01-2011:12$,

Models	Strategy	Risk Aversion									
		$\gamma=1$	$\gamma=2$	$\gamma=3$	$\gamma=4$	$\gamma=5$	$\gamma=6$	$\gamma=7$	$\gamma=8$	$\gamma=9$	$\gamma=10$
Comb.All	*no short-sale*	64.48	81.44	26.72	33.92	87.20	5.44	35.36	79.68	61.76	26.88
	short-sale	68.32	70.24	30.24	15.36	80.16	11.68	5.76	7.68	67.36	25.12
	(margin = 0.5)	57.60	69.44	2.24	15.36	63.68	11.68	1.12	7.20	57.92	22.40
Comb.SV	*no short-sale*	44.48	71.04	1.28	51.52	72.64	3.52	11.84	83.36	24.80	29.44
	short-sale	38.56	69.76	1.60	18.88	60.96	8.32	1.28	8.80	33.28	24.16
	(margin = 0.5)	34.24	66.08	1.60	18.88	39.84	8.32	0.96	8.32	14.40	23.20
Comb.EGARCH	*no short-sale*	56.16	89.44	1.28	90.24	75.52	5.60	14.56	93.12	11.20	32.80
	short-sale	41.76	78.72	1.92	29.60	55.20	7.20	1.76	11.36	13.60	26.56
	(margin = 0.5)	58.08	78.56	1.92	29.60	24.96	7.20	2.08	10.72	10.56	26.40
Comb.MF	*no short-sale*	81.92	99.84	16.00	91.84	47.68	12.64	74.24	95.36	25.12	1.28
	short-sale	70.40	99.84	11.84	65.28	20.16	16.48	5.12	12.80	22.40	1.28
	(margin = 0.5)	70.72	99.84	11.84	65.28	17.44	16.48	5.12	12.80	19.84	1.28
ATIC. Factor	*no short-sale*	71.68	99.84	5.76	98.56	72.64	12.96	47.04	96.32	20.96	1.28
	short-sale	55.88	99.84	4.00	65.28	38.08	19.20	3.20	13.76	16.64	1.28
	(margin = 0.5)	66.08	99.84	4.00	65.28	24.16	19.20	4.00	13.76	14.72	1.28

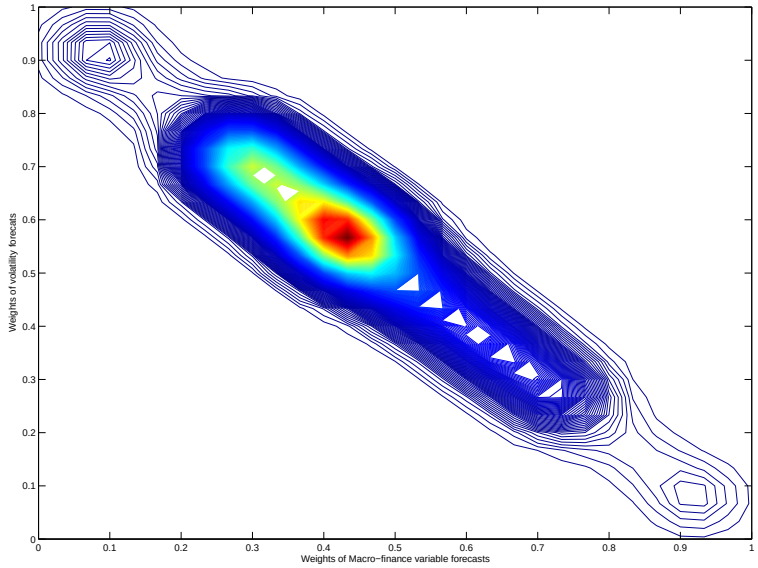
Note: *Table*.3.*b* treats each point in time as the entire sample to calculate CERs, and counts the time frequency when CERs yielded by other forecasts surpass those yielded by *Comb.RV*.

Figure.1.An Illustration of Forecasts of Combination



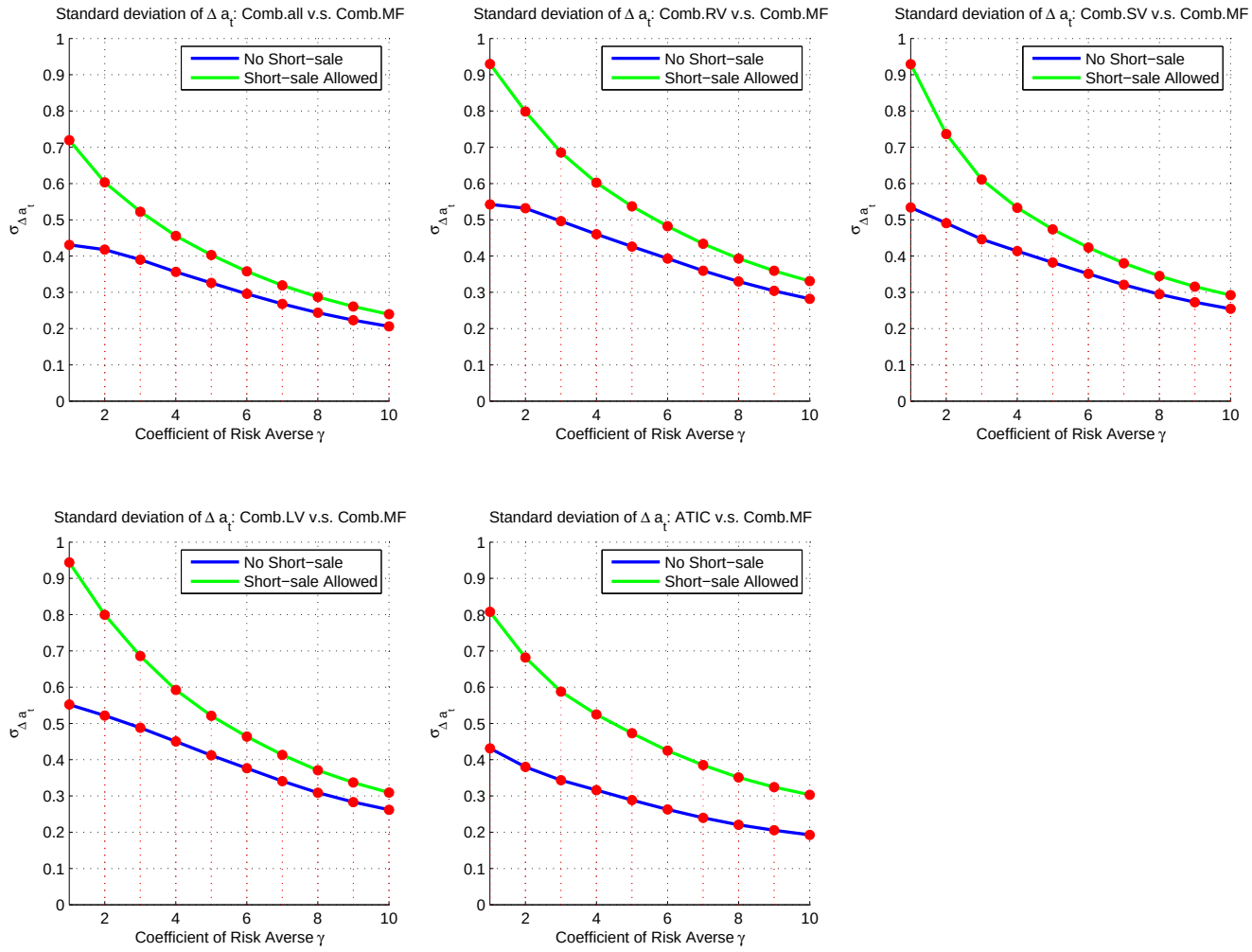
Note: The blue line represents the predictive distribution made by book-to-market ratio. The red line represents the predictive distribution made by term-spread. The black line is the combined predictive distribution produced by the optimal predictive pool using all 11 macro-finance variables.

Figure.2.Weights Allocation of the Optimal Prediction Pool



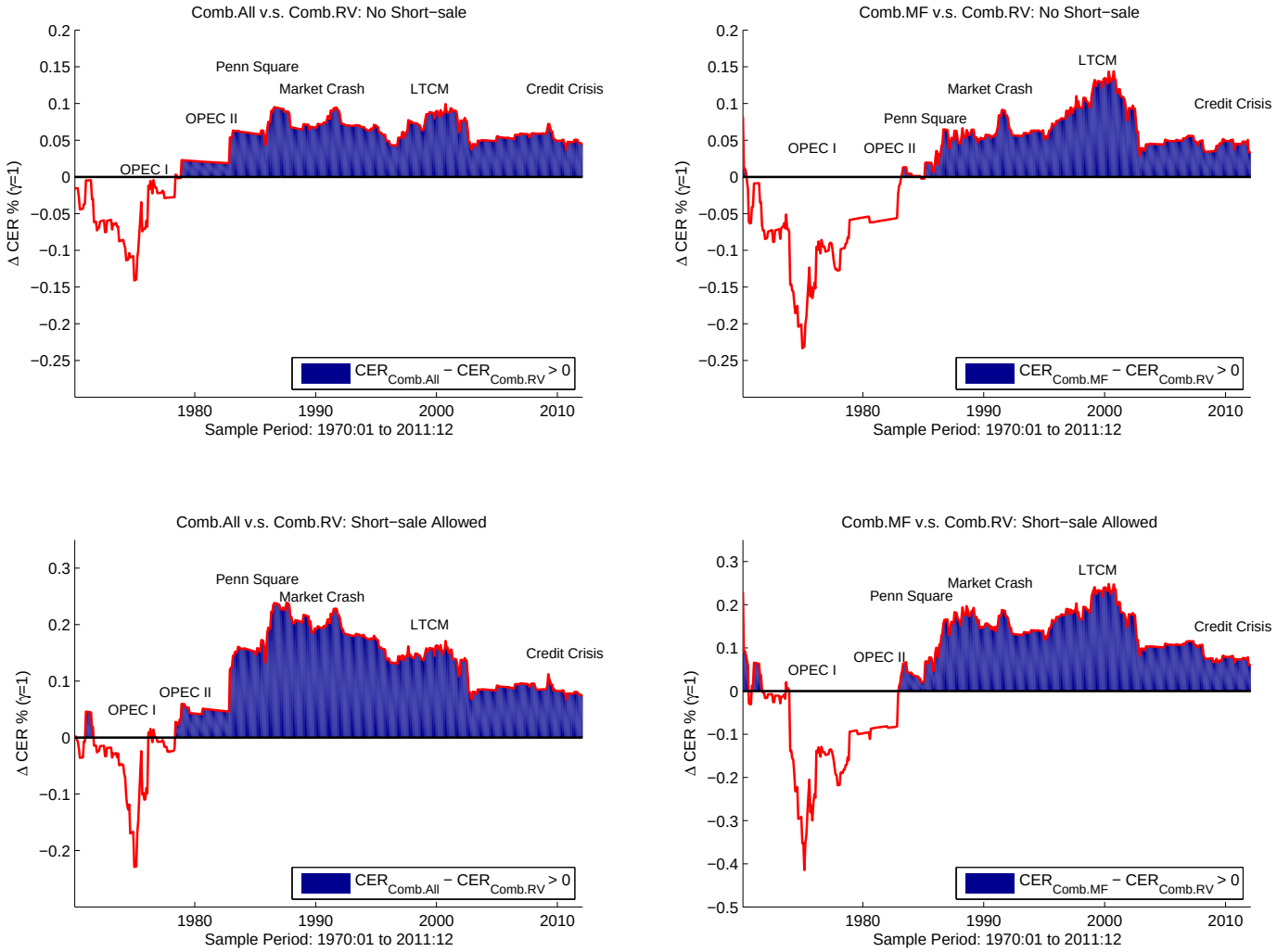
Note: The contour lines cluster at $(0.41, 0.59)$, which indicates that the macro-finance variables contribute 40% forecasting accuracy to the optimal pool, and the volatility models contribute 60% forecasting accuracy to the optimal pool.

Figure.3.1 The Convergence of the Portfolio Choice



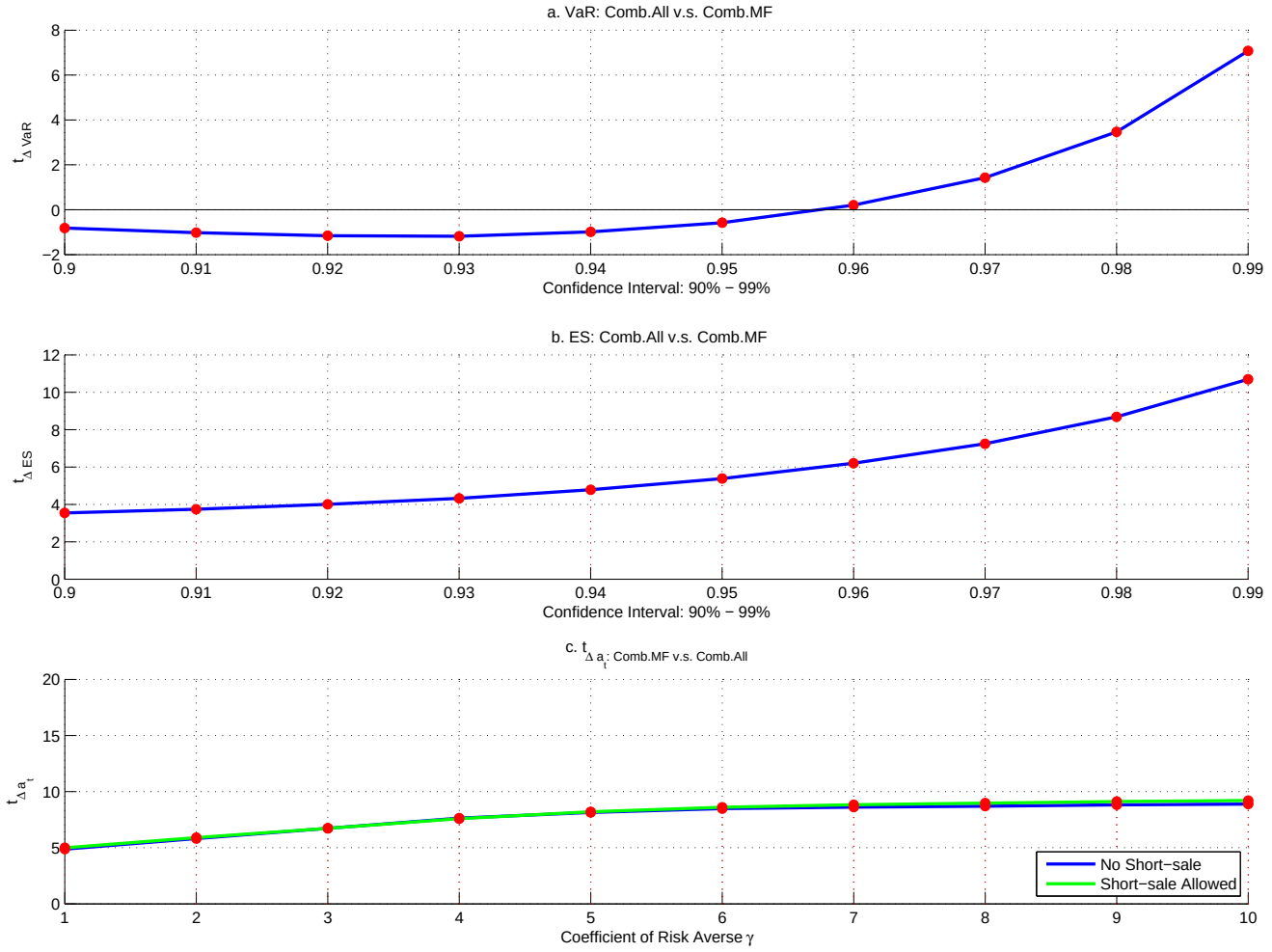
Note: *Figure.*3.1 shows that as the investors become more risk averse, the variance of the difference in portfolio choices declines, and the ratio allocated to risky assets converges, regardless of whether short-sales are allowed or not.

Figure.3.2. The Comparison of CER



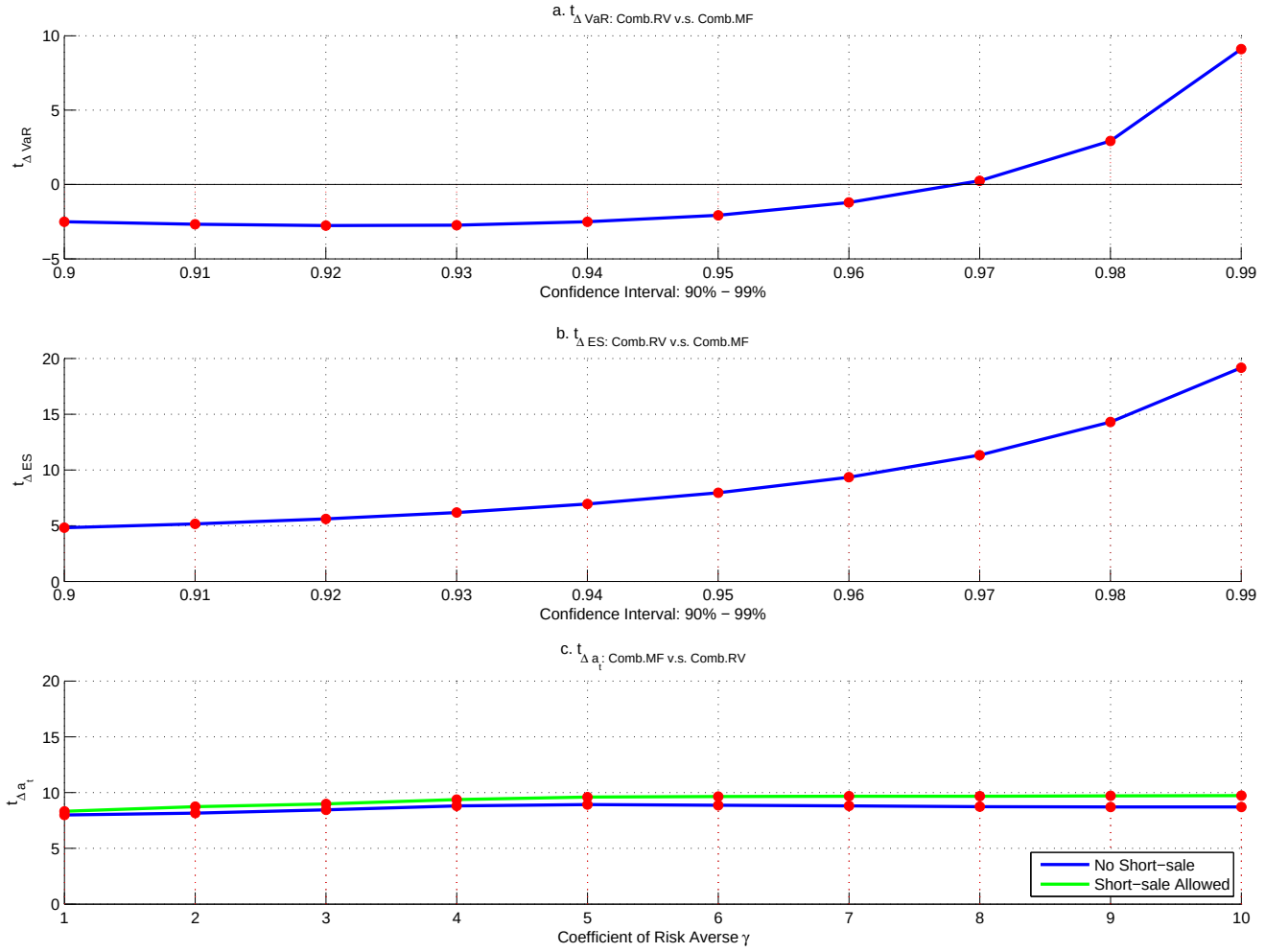
Note: The blue region refers to the time periods when *Comb.MF* or *Comb.All* outperforms *Comb.RV*, which mainly covers "the Great Moderation"

Figure.3.3 Risk Measure versus Portfolio Choice: Comb.All vs Comb.MF



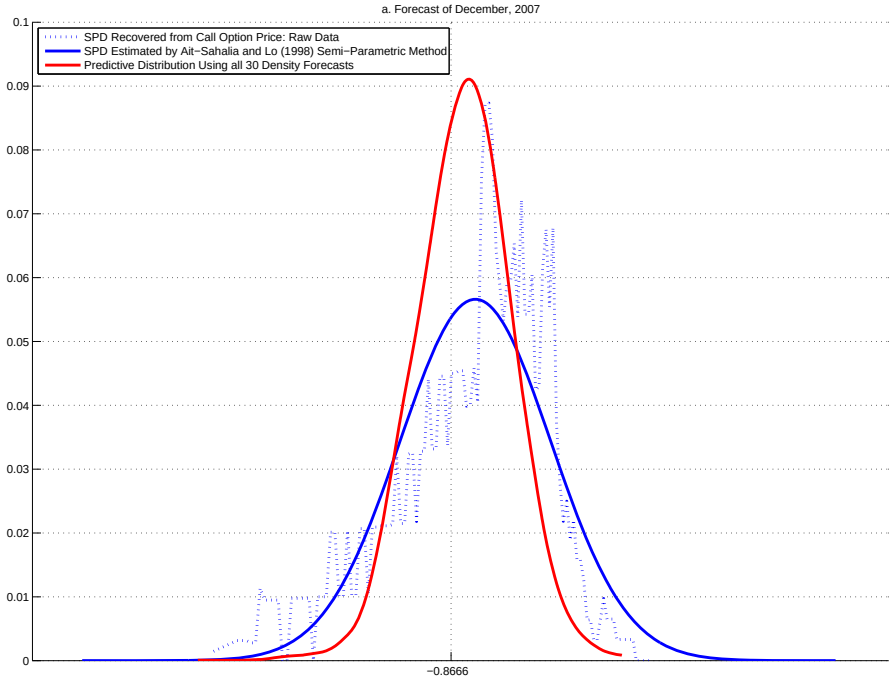
Note: *a*. and *b*. plot the test statistics that tests whether the value-at-risk and the expected short-fall formed by $Comb.All$ are equal to those formed by the $Comb.MF$. *c*. plot the test statistics that tests whether the fraction wealth invested in risky asset lead by $Comb.MF$ is significantly larger than $Comb.All$.

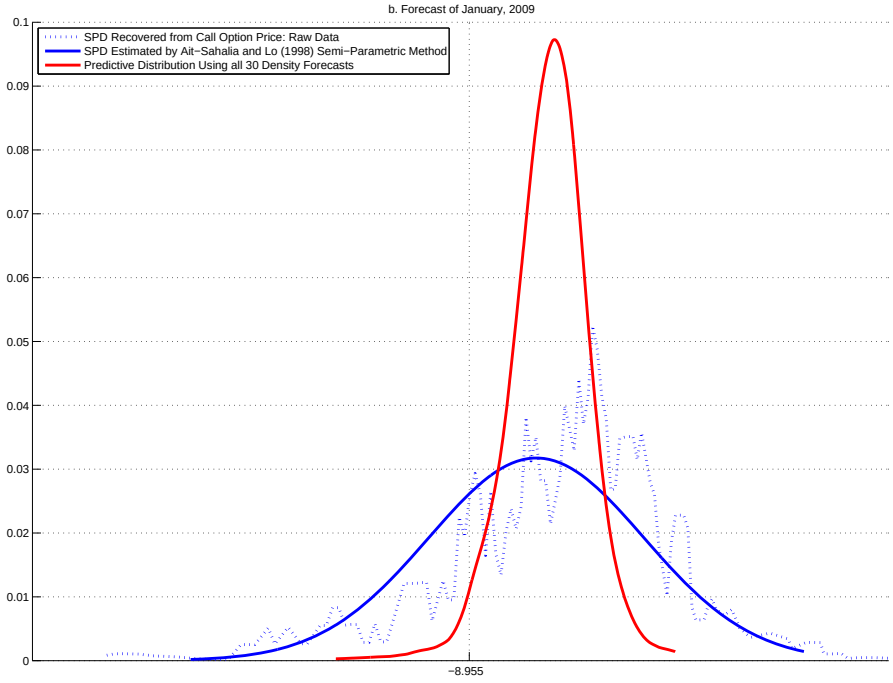
Figure.3.4 Risk Measure versus Portfolio Choice: Comb.RV vs Comb.MF



Note: *a*. and *b*. plot the test statistics that tests whether the value-at-risk and the expected short-fall formed by $Comb.RV$ are equal to those formed by the $Comb.MF$. *c*. plot the test statistics that tests whether the fraction wealth invested in risky asset lead by $Comb.MF$ is significantly larger than $Comb.RV$.

Figure.4.a-b State Price Density v.s. Density Forecast of *Comb.All*





References

Aıt-Sahalia, Y., & Lo, A. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *The Journal of Finance*, *53*, 499–547.

Aıt-Sahalia, Y., Wang, Y., & Yared, F. (2001). Do option markets correctly price the probabilities of movement of the underlying asset? *Journal of Econometrics*, *102*, 67–110.

Amisano, G., & Giacomini, R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business and Economic Statistics*, *25*, 177–190.

Ando, T., & Tsay, R. (2011). Quantile regression models with factor-augmented predictors and information criterion. *The Econometrics Journal*, *14*, 1–24.

Bai, J., & Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, *146*, 304–317.

Banz, R., & Miller, M. (1978). Prices for state-contingent claims: Some estimates and applications. *Journal of Business*, (pp. 653–672).

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics*, (pp. 542–547).

Breeden, D., & Litzenberger, R. (1978). Prices of state-contingent claims implicit in option prices. *Journal of business*, (pp. 621–651).

Cenesizoglu, T., & Timmermann, A. (2008). Is the distribution of stock returns predictable? *Working Paper*, .

Chernozhukov, V., Fernández-Val, I., & Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, *78*, 1093–1125.

Clark, T., & McCracken, M. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, *105*, 85–110.

Clark, T., & McCracken, M. (2005). Evaluating direct multi-step forecasts. *Econometric Reviews*, *24*, 369–404.

Clark, T., & McCracken, M. (2009). Tests of equal predictive ability with real-time data. *Journal of Business and Economic Statistics*, *27*, 441–454.

Diebold, F., & Mariano, R. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, *13*.

Diks, C., Panchenko, V., & van Dijk, D. (2011). Likelihood-based scoring rules for comparing density forecasts in tails. *Journal of Econometrics*, .

Durham, G., & Geweke, J. (2011). Improving asset price prediction when all models are false. *NBER Summer Institute, Working Group on Forecasting and Empirical Methods in Macroeconomics and Finance*, .

French, G. S., K.R., & Stambaugh, R. (1987). Expected Stock Returns and Volatility. *Journal of Financial Economics*, *19*, 3–29.

Gaglianone, W., & Lima, L. (2009). Constructing Density Forecasts from Quantile Regression. *Journal of Money, Credit, and Banking*, *44*, 1589–1607.

Gallant, A., Hsieh, D., & Tauchen, G. (1997). Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics*, *81*, 159–192.

Geweke, J., & Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, *forthcoming*.

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*, 1545–1578.

Jacquier, E., Polson, N., & Rossi, P. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, *122*, 185–212.

Kim, S., Shephard, N., & Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with arch models. *The Review of Economic Studies*, *65*, 361.

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, (pp. 33–50).

Manzan, S., & Zerom, D. (2009). Are Macroeconomic Variables Useful for Forecasting the Distribution of US Inflation? *NBER Summer Institute, Working Group on Forecasting and Empirical Methods in Macroeconomics and Finance*, .

McCracken, M. (2007). Asymptotics for out of sample tests of granger causality. *Journal of Econometrics*, *140*, 719–752.

Merton, R. (1980). On estimating the expected return on the market* 1:: An exploratory investigation. *Journal of Financial Economics*, *8*, 323–361.

Nelson, D. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica: Journal of the Econometric Society*, (pp. 347–370).

Nelson, D. (1992). Filtering and forecasting with misspecified arch models i:: Getting the right variance with the wrong model. *Journal of Econometrics*, *52*, 61–90.

Newey, W., & West, K. (1987). West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*, 703–708.

Omori, Y., Chib, S., Shephard, N., & Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, *140*, 425–449.

Ross, S. (1976). Options and efficiency. *The Quarterly Journal of Economics*, *90*, 75–89.

Stock, J., & Watson, M. (1998). Diffusion indexes.

Stock, J., & Watson, M. (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, *41*, 788–829.

Stock, J., & Watson, M. (2005). Implications of dynamic factor models for VAR analysis. *NBER working paper*, .

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, (pp. 307–333).

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

Zhao, Y. (2011). Real Time Density Forecasts of Output and Inflation via Quantile Regression. *Ph.D. Dissertation, The Jonhs Hopkins University*, *Chapter I*.