https://doi.org/10.20965/jaciii.2020.p0477

Paper:

Predictability of China's Stock Market Returns Based on Combination of Distribution Forecasting Models

Yanyun Yao*, **Xiutian Zheng**,†**, **and Huimin Wang***

*Department of Applied Statistics, Shaoxing University
900 Chengnan Avenue, Yuecheng District, Shaoxing, Zhejiang 312000, China
E-mail: nan_xiao@126.com
**Department of Economics, Hangzhou Normal University Qianjiang College
16 Xuelin Street, Xiasha University Town, Hangzhou, Zhejiang 310018, China
E-mail: zxt1369@163.com
†Corresponding author

[Received October 25, 2019; accepted January 12, 2020]

No consensus exists in the literature on whether stock prices can be predicted, with most existing studies employing point forecasting to predict returns. By contrast, this study adopts the new perspective of distribution forecasting to investigate the predictability of the stock market using the model combination strategy. Specifically, the Shanghai Composite Index and the Shenzhen Component Index are selected as research objects. Seven models – GARCH-norm, GARCH-sstd, EGARCH-sstd, EGARCH-sstd-M, one-component Beta-t-EGARCH, two-component Beta-t-EGARCH, and the EWMA-based nonparametric model – are employed to perform distribution forecasting of the returns. The results of out-of-sample forecasting evaluation show that none of the individual models is "qualified" in terms of predictive power. Therefore, three combinations of individual models were constructed: equal weight combination, log-likelihood score combination, and continuous ranked probability score combination. The latter two combinations were found to always have significant directional predictability and excess profitability, which indicates that the two combined models may be closer to the real data generation process; from the perspective of economic evaluation, they may have a predictive effect on the conditional return distribution in China's stock market.

Keywords: predictability, distribution forecasting, model combination, GARCH, nonparametric

1. Introduction

The predictability of capital market returns has been of great interest to academicians. Two major theories are related to return predictability: the efficient market hypothesis (EMH) of Fama [1] and the adaptive market hypothesis (AMH) of Lo [2]. [3] extends EMH's weak-form effectiveness test to a return predictability test, and considers that the unpredictability of the market is equivalent to "weak-form effectiveness." However, AMH believes that the predictability of market returns is determined by the changing market environment and will respond to changes in investor statistics, financial systems, and market conditions. Traditional finance believes that the returns on capital markets are unpredictable; however, several studies have found predictable returns. Cochrane [4] even calls the predictability of market returns "new facts" in finance. Some studies, such as [5] and [6], believe that the predictability of the stock market is consistent with the AMH theory, and that there are periods of predictable returns. The research on the predictability of capital market returns is of vital importance to the development of financial theory. Cochrane [7] believes that the existence of predictability will modify all previous results based on the assumption of random walks in the stock market. Ferson et al. [8] points out that predictive variables found in the stock market through regression techniques will play a role in conditional pricing models. In addition, the out-of-sample predictability of financial returns has important implications for many areas such as asset pricing and portfolio allocation. Campbell and Thompson [9] find, through actual data, that even weak predictability can improve the effect of asset allocation.

The non-linear and non-stationary characteristics of the stock market make it a complex system, and its complexity is further related to various factors such as political events, market news, quarterly earnings reports, international influences, and investor trading behaviors. Therefore, forecasting stock market returns is a difficult task. The commonly used methods in the literature are statistical prediction models under a certain error evaluation criterion, such as ARMA time-series models and grey prediction models. In recent years, machine learning approaches have also become popular; see [10]. Affected by changes in the market environment, parameter instability of the model exists objectively; as a result, time-varying parameter models are used in empirical studies such as [11] and [12]. Model uncertainty is also an important factor affecting prediction performance. To reduce model uncertainty, combination strategies have been used, such as in [13] and [14]. However, existing research on the predictability of the stock market mainly relies on point prediction. Point forecasting means that information is released in a "centralized trend" model, without considering uncertainty or related risks; governments and investors, however, are more concerned about forecasted risk measures, which means that forecasts should include risk management. Zheng et al. [15] point out that the predictability of returns has changed from conditional means predictability to direction predictability, and direction predictability can come not only from the sequence dependence of the conditional mean, but also from other higher-order sequence dependence. Christoffersen and Diebold [16] find that the sequence-dependent characteristics of volatility can be used to predict the direction of the market. In theory, point prediction, risk prediction, and direction prediction can all be included in distribution prediction. Distribution prediction estimates the future conditional distribution function of a random variable based on existing information, and can fully describe the uncertainty characteristics of the predictor. Thus, this article provides new ideas for studying the predictability of the stock market from the perspective of out-of-sample distribution prediction, namely, distribution forecasting, which could lead to different conclusions.

China's stock markets, in particular, have witnessed significant sudden uncertainties, causing stock prices to rise and plunge, which has led to the obvious peak and thick-tail characteristics of its stock market distribution. Li et al. [17] explained the microeconomic basis for the time variation of return distribution, and confirmed the existence of time variation through an empirical study of the return data of a complete bull-bearing market cycle of the Shanghai Composite Index. Therefore, the one-step-ahead rolling window method is used to forecast distribution in this paper. Parametric methods are most frequently used for distribution forecasting in the literature. In particular, the GARCH family model, a simple and popular parametric model, has been widely used in the modeling and prediction of economic and financial time series. The model uses a two-step method to construct the distribution forecasting. First, it models the conditional dynamics of the mean and the conditional volatility of the studied time series. The mean equation uses a point forecasting model such as the ARMA model or the neural network model, while the volatility equation is set to a GARCH process such as the GJR-GARCH or EGARCH process. Second, it sets the residual distribution as a parameter distribution, such as normal distribution or Student t-distribution. After the parameters are estimated, the distribution of the residuals is simulated to obtain the predicted distribution.

To improve the explanation power of the GARCH model, the GARCH-norm, GARCH-sstd, EGARCH-sstd, and EGARCH-sstd-M models are employed, considering the setting of residual distribution, leverage effect, and risk compensation effect, and so on. The Beta-t-EGARCH model, which is robust to sudden uncertainty modeling, is employed as well, including both one-component and two-component models. The EWMA-based nonparametric model, which gains a good fitting effect in Yao and Xu [18], is also included. Due to the different focuses of the different models and the existence of model uncertainty, this study will also examine the strategy of combined models under the guidance of the marginal calibration figure.

Distribution is a complete description of the random variable. With the prediction of the return distribution, the corresponding direction forecasting and point forecasting can also be derived. Compared with statistical predictability, market investors are more concerned about the profitability of making investment decisions using return forecasts [19]. Therefore, the evaluation of forecasting performance is carried out from the perspective of economic predictability. Two test statistics are employed. One is the directional accuracy test of Pesaran and Timmermann (the PT test) [20], and the other is the excess profit test of Anatolyev and Gerko (the AG test) [21]. To investigate the excess profitability of the return forecast, a simulated trading strategy is designed for the mean and median forecasts generated from the forecasted distribution of returns.

2. Models and Methods

As mentioned in [22], investors face real-time decision-making problems; therefore, the following individual models are considered from a time-varying perspective, that is, the model type is unchanged, but the preset parameters may change over time. Seven individual models will be detailed below, followed by an explanation of how distribution forecasts are obtained.

2.1. Models

Given the sequence of return variables for financial assets, a large number of empirical studies have shown that their distributions have the following characteristics: (i) a heavy tail, that is, a tail that is heavier than normal distribution; (ii) volatility clustering, that is, large fluctuations that tend to follow large fluctuations, and alternating periods of gradual change and large fluctuations; and (iii) an accumulation of the Gaussian property. When the sampling frequency is reduced, the central limit theorem is established, and the return over a long period of time tends to be normally distributed. The ARCH model proposed by [23] and the GARCH model of [24] can capture the above three characteristics.

2.1.1. GARCH-norm Model

In the financial volatility modeling, the GARCH $(1,1)$ model with normal distribution residuals is often chosen as the baseline model, specifically defined as M_1, that is:

$$M_1 : \begin{cases} Y_t = \mu + \varepsilon_t, & \varepsilon_t | I_{t-1} \overset{iid}{\sim} N(0, h_t), \\ h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}. \end{cases} \quad . \quad (1)$$

2.1.2. GARCH-sstd Model

Wang and Wang [25] argue that the skewed-t distribution provides the best choice for the characterization and prediction of the actual volatility characteristics of China's stock markets. Therefore, GARCH $(1,1)$ with skewed-t distribution residuals, named the GARCH-sstd model, is employed. It is defined as:

$$M_2: \begin{cases} Y_t = \mu + \varepsilon_t, \\ h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}, \\ \varepsilon_t | I_{t-1} \overset{iid}{\sim} \\ \qquad \dfrac{2}{(\lambda+\lambda^{-1})\sqrt{h_t}} \left[f_v\left(\dfrac{\varepsilon_t}{\lambda\sqrt{h_t}}\right) I(\varepsilon_t \geq 0) \right. \\ \qquad \left. + f_v\left(\dfrac{\lambda\varepsilon_t}{\sqrt{h_t}}\right) I(\varepsilon_t < 0) \right], \end{cases} \tag{2}$$

where $I(\cdot)$ is a indicator function, $f_v(\cdot)$ represents the density function of Student t-distribution with the freedom degree v, which is

$$f_v(x) = \frac{\sqrt{\dfrac{v}{v-2}} \quad \Gamma\left(\dfrac{v+1}{2}\right)}{\sqrt{\pi} \quad \Gamma\left(\dfrac{v}{2}\right)} \left(1 + \frac{x^2}{v-2}\right)^{-\frac{v+1}{2}}. \tag{3}$$

2.1.3. EGARCH-sstd

Many studies have shown that the negative skew of stock returns is because traders react more strongly to unfavorable information than to favorable information. Therefore, many scholars believe that the leverage effect of returns should be considered. The relevant models are the TARCH model (or GJR-GARCH model) and the EGARCH model. The EGARCH model introduced by [26] is widely used. The EGARCH $(1,1)$ model with skewed-t distribution residuals is denoted as the EGARCH-sstd model and it is defined as:

$$M_3: \begin{cases} Y_t = \mu + \varepsilon_t, \\ \ln h_t = \omega + \beta \ln h_{t-1} + \alpha\left|\dfrac{\varepsilon_{t-1}}{\sqrt{h_t}}\right| + \gamma\dfrac{\varepsilon_{t-1}}{\sqrt{h_t}}, \\ \varepsilon_t | I_{t-1} \overset{iid}{\sim} \\ \qquad \dfrac{2}{(\lambda+\lambda^{-1})\sqrt{h_t}} \left[f_v\left(\dfrac{\varepsilon_t}{\lambda\sqrt{h_t}}\right) I(\varepsilon_t \geq 0) \right. \\ \qquad \left. + f_v\left(\dfrac{\lambda\varepsilon_t}{\sqrt{h_t}}\right) I(\varepsilon_t < 0) \right]. \end{cases} \tag{4}$$

2.1.4. EGARCH-sstd-M Model

Financial theory suggests that assets with higher observable risks can achieve higher average returns because people generally believe that the return on financial assets should be proportional to their risk. The greater the risk, the higher the expected return. The ARCH model, which includes conditional variance (or standard deviation) to represent expected risk, is called the ARCH-M model and was introduced by [27]. In this paper, the conditional standard deviation is used to reflect the conditional risk and produce the EGARCH-sstd model, obtaining the EGARCH-sstd-M model, that is:

$$M_4: \begin{cases} Y_t = \mu + \delta\sqrt{h_t} + \varepsilon_t, \\ \ln h_t = \omega + \beta \ln h_{t-1} + \alpha\left|\dfrac{\varepsilon_{t-1}}{\sqrt{h_t}}\right| + \gamma\dfrac{\varepsilon_{t-1}}{\sqrt{h_t}}, \\ \varepsilon_t | I_{t-1} \overset{iid}{\sim} \\ \qquad \dfrac{2}{(\lambda+\lambda^{-1})\sqrt{h_t}} \left[f_v\left(\dfrac{\varepsilon_t}{\lambda\sqrt{h_t}}\right) I(\varepsilon_t \geq 0) \right. \\ \qquad \left. + f_v\left(\dfrac{\lambda\varepsilon_t}{\sqrt{h_t}}\right) I(\varepsilon_t < 0) \right]. \end{cases} \tag{5}$$

2.1.5. One-Component Beta-Skew-t-EGARCH Model

The Beta-t-EGARCH model was originally proposed by [28] and [29]. Harvey and Sucarrat [30] extended it to the skewed case, producing the Beta-Skew-t-EGARCH model. Blazsek and Villatoro [31] and Blazsek and Mendoza [32] have shown that compared with other GARCH models, this model exhibits strong robustness against jumps or outliers and, empirically, can effectively capture the sudden uncertainty of financial returns. The one-component Beta-Skew-t-EGARCH model is defined as:

$$M_5: \begin{cases} Y_t = \mu + e^{\lambda_t}\varepsilon_t = \mu + \sqrt{h_t}\varepsilon_t, \\ \varepsilon_t | I_{t-1} \overset{iid}{\sim} sstd(0, \sigma_\varepsilon^2, \nu, \gamma), \\ \lambda_t = \omega + \lambda_t^\dagger, \\ \lambda_t^\dagger = \phi_1\lambda_{t-1}^\dagger + \kappa_1 u_{t-1} \\ \qquad + \kappa^* sgn(-Y_{t-1})(u_{t-1}+1). \end{cases} \tag{6}$$

To meet the stability conditions, additional restrictions are required: $|\phi_1| < 1, \nu > 2, \gamma > 0$. u_t is called conditional score, which represents the derivative of the log likelihood score of Y_t to λ_t at time t, that is:

$$\begin{aligned} u_t &= \frac{\partial ln f_{Y_t}(y_t)}{\partial\lambda_t} \\ &= \frac{(\nu+1)(y_t^2 + y_t\mu_\varepsilon e^{\lambda_t})^2}{\nu e^{2\lambda_t}\gamma^{2sgn(y_t+\mu_\varepsilon e^{\lambda_t})} + (y_t+\mu_\varepsilon e^{\lambda_t})^2} - 1, \end{aligned} \tag{7}$$

where $sgn(\cdot)$ represents a symbolic function. As described by [30], u_t can also be conveniently written as:

$$u_t = \frac{(\nu+1)(\varepsilon_t^{*2} - \mu_\varepsilon\varepsilon_t^*)}{\nu\gamma^{2sgn(\varepsilon_t^*)} + \varepsilon_t^{*2}} - 1, \tag{8}$$

where ε_t^* is a non-centralized skewed t distribution variable. When the conditional distribution is symmetrical, i.e., $\gamma = 1$, then:

$$\frac{u_t+1}{\nu+1} \sim Beta\left(\frac{1}{2}, \frac{\nu}{2}\right). \tag{9}$$

This is why the model is called as Beta-t-EGARCH model.

2.1.6. Two-Component Beta-Skew-t-EGARCH Model

As stated in [33], the square of financial returns tends to exhibit long-term memory characteristics. The two-component model of volatility can be adapted to the long-memory nature by decomposing volatility into a long-term component and a short-term component. The transient changes produced by shocks are only reflected in the short-term component. The two-component Beta-Skew-t-EGARCH model proposed by [30] is represented as:

$$M_6 : \begin{cases} Y_t = \mu + e^{\lambda_t}\varepsilon_t = \mu + \sqrt{h_t}\varepsilon_t, \\ \varepsilon_t | I_{t-1} \overset{iid}{\sim} sstd(0, \sigma_\varepsilon^2, \nu, \gamma), \\ \lambda_t = \omega + \lambda_{1,t}^\dagger + \lambda_{2,t}^\dagger, \\ \lambda_{1,t}^\dagger = \phi_1 \lambda_{1,t-1}^\dagger + \kappa_1 u_{t-1}, \\ \lambda_{2,t}^\dagger = \phi_2 \lambda_{2,t-1}^\dagger + \kappa_2 u_{t-1} \\ \qquad + \kappa^* sgn(-Y_{t-1})(u_{t-1}+1), \end{cases} \quad (10)$$

where $\lambda_{1,t}^\dagger$ and $\lambda_{2,t}^\dagger$ represent the long-term component and the short-term component respectively, the definition of other parameters is the same as in M_5. Here the leverage effect only displays in the short-term fluctuation. To make the model identifiable, the restrictions $\nu > 2, \gamma > 0, |\phi_1| < 1, |\phi_2| < 1$ and $\phi_1 \neq \phi_2$ are needed.

2.1.7. NP-EWMA Model

The NP-EWMA model refers to the EWMA-based nonparametric model proposed by [34]. [18] shows that from the perspective of quantile evaluation, the model has a better in-sample fitting effect than the GARCH-norm model, and has a certain typical fact interpretation ability. Like [18], the kernel method is used to investigate the modeling of the overall condition distribution of returns. Contrary to [18], the out-of-sample forecasting is considered here rather than an in-sample fit. It is recorded as model M_7, specially,

$$M_7 : F_{t+1}(y) = \omega F_t(y) + (1-\omega) H\left(\frac{y - y_t}{h}\right), \quad (11)$$

where $F_t(y)$ represents the cumulative distribution function of return at time t, $H(\cdot)$ is the kernel function with a cumulative distribution form, ω is a decay factor and h is the bandwidth.

2.2. The Estimation and Return Distribution Forecasting of Individual Models

Elliott and Timmermann [35] propose the full-parameter method of density estimation, which assumes that the conditional density $p_Y(y|Z,\theta)$ is known except the parameters θ. The classical method uses data Z to obtain parameter estimates $\hat{\theta}(Z)$, and substitutes these estimates to obtain the density forecasts $p_Y(y|Z,\hat{\theta}(Z))$, at the same time as the cumulative distribution function $F_Y(y|Z,\hat{\theta}(Z))$.

To achieve the estimation $\hat{\theta}(Z)$, some form of loss function needs to be assumed to obtain an estimate by maximizing or minimizing these specific forms of loss functions. The most common loss function focuses on the degree of "proximity" between the candidate density $p(y|Z,\theta)$ and the true density "$p_0(y|Z,\theta_0)$," which is the Kullback-Leibler (KL) distance. The KL distance between the true distribution p_0 and the parameter distribution p is:

$$\begin{aligned} KL(p_0, p) &= E_{p_0}\left(ln\frac{p_0}{p}\right) \\ &= E_{p_0}[ln(p_0(y|Z,\theta_0))] - E_{p_0}[ln(p(y|Z,\theta))]. \end{aligned} \quad (12)$$

Obviously, minimizing the KL distance is equivalent to maximizing $E_{p_0}[ln(p(y|Z,\theta))]$. $E_{p_0}[ln(p(y|Z,\theta))]$ is the expectation of the log likelihood function on θ, also known as the expected logarithmic score. Therefore, in terms of density (or distribution) forecasting, the methods of Maximum Likelihood Estimation (MLE) and minimizing the KL distance are the same.

As to the model M_1, $Y_t \sim N(0,h_t)$, then the density function of Y_t is:

$$f_{Y_t}(y) = \frac{1}{\sqrt{2\pi h_t}} e^{-\frac{(y-\mu)^2}{2h_t}}, \quad (13)$$

so the log likelihood contribution of Y_t at time t is:

$$l_t = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln h_t - \frac{1}{2}\frac{(y_t-\mu)^2}{h_t}, \quad (14)$$

hence:

$$\begin{aligned} E_{p_0}[\ln(p(y|Z,\theta))] &= \frac{1}{T}\sum_{t=1}^{T} l_t \\ &= -\frac{1}{2T}\left[T\ln(2\pi) + \sum_{t=1}^{T}\ln h_t + \sum_{t=1}^{T}\frac{(y_t-\mu)^2}{h_t}\right]. \end{aligned} \quad (15)$$

MLEs of the parameters in M_1 are achieved by maximizing Eq. (15).

As to models M_2, M_3, the skewed-t distribution is used for the residual distribution, so the distribution of Y_t is

$$\begin{aligned} F_{Y_t}(y) &= P(Y_t \le y) = P(\mu + \sqrt{h_t} z_t \le y) \\ &= P\left(z_t \le \frac{y-\mu}{\sqrt{h_t}}\right) = F_{z_t}\left(\frac{y-\mu}{\sqrt{h_t}}\right), \end{aligned} \quad (16)$$

where $z_t = \varepsilon_t/\sqrt{h_t}$ is the standardized residual, with the mean 0 and the variance 1, $F_{z_t}(\cdot)$ is the cumulative distribution of z_t, thus the density of Y_t is:

$$f_{Y_t}(y) = \frac{1}{\sqrt{h_t}} f_{z_t}\left(\frac{y-\mu}{\sqrt{h_t}}\right). \quad (17)$$

Furthermore, similar to the processing of M_1, the parametric estimates of M_2 and M_3 can be obtained by maximizing $E_{p_0}[\ln(p(y|Z,\theta))$.

As to the models M_4, M_5, M_6, it is only necessary to use the conditional mean equation for refinement. Similarly, the parameter estimates can be obtained by MLE.

Thus, as to $M_1, M_2, \ldots, M_6$, after obtaining the parameter estimation, the distribution forecasting is available, which is:

$$\hat{F}_{Y_{t+1|t}}(y) = \hat{F}_{z_{t+1|t}}\left(\frac{y-\hat{\mu}}{\sqrt{\hat{h}_{t+1}}}\right). \quad (18)$$

As to M_7, the initial m samples are employed to generate the initial distribution $F_1(y)$ according to [34]. Here we set $m = 100$. Thus,

Table 1. P value of the Berkowitz test.

	SHCI			SZCI		
Model	Whole	5% tail	1% tail	Whole	5% tail	1% tail
M_1	0.2143	**<0.0001**	**<0.0001**	0.3836	**<0.0001**	**<0.0001**
M_2	0.1135	0.5876	**0.0430**	0.2585	0.2035	0.2324
M_3	0.2483	0.3214	0.0641	0.5028	0.0843	0.1311
M_4	0.1716	0.2546	0.1417	0.3438	0.0895	0.1632
M_5	**<0.0001**	**0.0298**	0.4404	**<0.0001**	0.1047	0.7620
M_6	**<0.0001**	**0.0128**	0.1775	**<0.0001**	0.0701	0.6645
M_7	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**

$$E_{p_0}[\ln(p(y|Z,\theta)) = \frac{1}{T-m}\sum_{t=m}^{T-1}\ln f_{t+1|t}\hat{(}y_{t+1}) = \frac{1}{T-m}\sum_{t=m}^{T-1}\ln\left[\frac{1}{h}\sum_{i=1}^{t}K\left(\frac{y_{t+1}-y_i}{h}\right)W_{t,i}(\omega)\right], \quad (19)$$

where $K(\cdot)$ is a kernel function with a density form, here it is Gaussian. According to [36], the weight function $W_{t,i}$ is set as:

$$W_{t,i} = \frac{1-\omega}{1-\omega^t}\omega^{t-i}. \quad (20)$$

After obtaining the parameter estimates of ω and h, the distribution forecast of M_7 is:

$$\hat{F}_{Y_{t+1|t}}(y) = \hat{\omega}\hat{F}_{t|t-1}(y) + (1-\omega)H\left(\frac{y-y_t}{\hat{h}}\right). \quad (21)$$

3. Empirical Study

3.1. Data Description

In this study, the Shanghai Composite Index (SHCI) and the Shenzhen Component Index (SZCI) are selected as research objects. The time span is from January 2, 2004 to December 30, 2016, with a total of 13 years of transaction data. The return is calculated as a percentage of the logarithmic yield, that is:

$$Y_t = 100 \times (\ln p_t - \ln p_{t-1}), \quad (22)$$

where p_t is the close price at time t. The sample size is $T = 3158$. As shown in [18], the returns of SHCI and SZCI have similar volatility characteristics. The violent ups and downs are frequent, the volatility is clustering and the ARCH effect is significant.

Rolling samples are used to estimate the parameters and forecast the distribution of the returns, the fitting sample size is $T_0 = 2425$, the period of first estimation corresponds to January 2, 2004 to December 31, 2013, and the forecasting sample size is $n = T - T_0 = 733$.

3.2. Statistical Evaluation of Distribution Forecasts

3.2.1. PIT Evaluation

Nowotarski and Weron [37] point out that according to the preamble principle of [38], the distribution forecasting evaluation only needs to perform the pairwise performance of the forecasting distribution and the realized value, i.e., $(F_{t|t-1}, y_t)$. Dawid [38] and Diebold et al. [39] propose to use Probability Integral Transform (PIT). Berkowitz [40] converts the PIT into a normal distribution. In addition, Ghalanos [41] points out that the Berkowitz test also provides a tail test based on a truncated normal distribution. The null hypothesis is that the normalized tail data has a mean of 0 and a variance of one. Hong and Li [42] refer to PIT as generalized residual, and based on [43], a nonparametric omnibus test is proposed. HL test can simultaneously test identically distributed random variables (IID) and $U[0,1]$. Thus, the Berkowitz test and HL test are performed in this paper.

The Berkowitz test is performed on the PIT sequences, from three types respectively: the whole distribution, the 5% tail, and the 1% tail. The P value of the test is shown in **Table 1**, and the significant value at 5% level is displayed in bold.

From **Table 1**, M_3 and M_4 for SHCI and SZCI with M_2 for SZCI fail to reject the null hypothesis, which is considered that the PITs obey IID $U[0,1]$; M_1 does not reject the null hypothesis H_0 from the whole distribution, but rejects H_0 for the tails, while M_5 and M_6 both reject H_0 for the whole distribution but do not reject H_0 for the 1% tail. M_7 seems to be the worst one, and all three tests reject the null hypothesis significantly.

Further, the statistic results of HL test are shown in **Table 2**. The critical value of HL test at 5% level is 1.6445, the significant values (>1.6445) are emphasized in bold. According to the W statistic, all the PIT sequences reject the null hypothesis. However, from the relative size of W statistic, the models: M_2, M_3, and M_4 are significantly better than other models. Use $>$ to represent "is better than," then we can get that $M_3 > M_2 > M_4$ for SHCI and $M_4 > M_3 > M_2$ for SZCI. According to the $M(i,j)$ statistics, the forecasting performance of SZCI is better than that of SHCI for all models, only $M(4,4)$ of SZCI for M_7 reject the null hypothesis. It is worth mentioning that for $M(1,2)$ of SHCI, only the nonparametric model M_7 does not reject the null, which implies that it is effective for modeling the influence of the second moment on the first moment. From PIT evaluation, it can be concluded that EGARCH-sstd (M_3) and EGARCH-sstd-M (M_4) are relatively better than other models for the given samples. EGARCH-sstd and EGARCH-sstd-M are both of residu-

Table 2. Statistics of the HL test.

Index	Model	M(1,1)	M(2,2)	M(3,3)	M(4,4)	M(1,2)	M(2,1)	W
SHCI	M_1	0.9881	**2.3171**	**4.0204**	**5.0232**	**5.0477**	−0.2937	**27.2216**
	M_2	0.1031	**1.7596**	**3.7017**	**5.0224**	**3.9549**	−0.7258	**8.3930**
	M_3	−0.5612	0.5300	**1.7098**	**2.3341**	**2.6123**	−1.2473	**7.6999**
	M_4	−0.4528	0.2735	1.2683	**1.8425**	**2.4420**	−1.2653	**8.7487**
	M_5	−0.2583	−0.1601	0.5022	0.8807	**1.9689**	−1.4915	**38.664**
	M_6	0.3777	1.2616	**3.1638**	**5.0018**	**3.2661**	−0.9842	**41.4306**
	M_7	−1.0435	0.7069	**4.2543**	**7.6053**	0.5634	−1.3167	**36.1231**
SZCI	M_1	−0.6151	−0.8919	−0.4603	0.1098	1.6088	1.4292	**20.7278**
	M_2	−1.1497	−1.1871	−0.8793	−0.5250	0.7277	−1.3447	**7.3893**
	M_3	−1.4604	−1.4671	−1.3898	−1.3219	−0.0254	−1.2099	**5.9005**
	M_4	−1.3712	−1.5429	−1.6324	−1.7193	−0.2495	−1.1385	**5.1786**
	M_5	−1.0691	−1.4588	−1.3900	−1.3445	0.2781	−1.5529	**21.0483**
	M_6	−0.6608	−1.1809	−0.6767	0.2743	1.2988	−1.8940	**20.8097**
	M_7	−1.8199	−0.9490	1.3478	**4.0676**	0.3184	−1.4521	**25.4860**

als with skewed-*t* distribution and examining the leverage effect. EGARCH-sstd-M also examines the risk compensation effect of returns.

3.2.2. Logarithmic Score

The logarithmic score is the negative of the logarithm of putting the observations into the predicted density (see [44]). Many academic works consider this scoring rule to be correct and to have many desirable attributes. In order to reflect the characteristic of the higher the score then the better the model, the logarithmic score here is defined as:

$$S_t = \ln \hat{f}_{t|t-1}(y_t). \qquad (23)$$

We call the model with the highest logarithmic score the Bayesian winner, and the *i*-th model becomes the Bayesian winner at time *t* means:

$$S_t^{(i)} > S_t^{(j)}, \ (\forall j \neq i). \qquad (24)$$

Obviously, the more the model becomes the Bayesian winner, the better the forecasting performance of the model. Hence, the times of the *i*-th model becomes Bayesian winner are counted as:

$$B_i = \#\{S_t^{(i)} > S_t^{(j)}\}, (\forall j \neq i), \qquad (25)$$

where $\#\{\cdot\}$ represents the counting operator.

The number and ranking of each model as the Bayesian winner are listed in the third and fourth columns of **Table 3**. The situations of SHCI and SZCI are similar, and only the orders of M_2 and M_6 are slightly different. Surprisingly, M_7, which is quite unsatisfactory in PIT evaluation, is far ahead, followed by M_4, then M_1.

The average log-scores and ranking are shown in the fifth and sixth columns of **Table 3**, and the results are consistent with the PIT evaluation. For the individual models, the rankings of SHCI and SZCI are identical, and M_2, M_3, and M_4 are the top three with the order: $M_3 > M_4 > M_2$.

3.2.3. Calibration, Sharpness, and CRPS

Marginal calibration, sharpness, and CRPS provided by [44] are a set of powerful distribution forecasting evaluation tools; they are nonparametric evaluation methods that do not rely on nested models. The marginal calibration charts of SHCI and SZCI are shown in **Fig. 1**. As a whole, the calibration of SZCI is better than that of SHCI, with the tail, especially the right tail, better calibrated than the middle part. M_2, M_3, and M_4 have better middle part calibration, and M_5 and M_6 have the best calibration at the tail. M_1 and M_7 seem to be underperforming, and M_1 is different from other models in the return interval $(0.5, 4)$, M_7 is also different from other models in the return range $(-3.5, -1.5)$ of SHCI, which leads us to combine the distribution forecasts to improve the calibration. Next, the sharpness is examined, and the average widths of the 50% and 90% confidence intervals are used for characterization, referring to [44]. The smaller the values, the better the sharpness. Confidence intervals are generated based on quantiles. The 50% confidence interval is $[F^{-1}(0.25), F^{-1}(0.75)]$ and the 90% confidence interval is $[F^{-1}(0.05), F^{-1}(0.95)]$. The results are shown in **Table 4**. Surprisingly, M_5, M_6, and M_7 have better sharpness. For SHCI, M_7 has a significantly smaller average width. For SZCI, M_5, M_6 and M_7 have significantly smaller average widths than M_1, M_2, M_3, and M_4. Finally, the CRPS scores and ranking of each model are calculated and listed in Columns 7 and 8 of **Table 3**. The difference between CRPS and the logarithmic score is that CRPS takes into account the sharpness and has better robustness generally. The smaller the CRPS, the better the model. It can be seen from **Table 3** that, like the logarithmic score, M_2, M_3 and M_4 are ranked in the top three, and M_3 is still the first, but under the CRPS criterion, M_2 is better than M_4. Compared to the logarithmic score evaluation, the ranking of M_1 has improved, from 6th to 4th, and the nonparametric model M_7 is superior to the Beta-t-EGARCH models M_5 and M_6.

3.3. Combination and Economic Evaluation

3.3.1. Combination Strategy of Models

Each model has its own advantages and disadvantages. In PIT evaluation, M_3 and M_4 are relatively superior. In

Table 3. Score and ranking of SHCI and SZCI.

Index	Model	Bayesian winner	Winner ranking	Average log-score	Log-score ranking	CRPS	CRPS ranking
SHCI	M_1	146	3	−1.7572	6	0.8557	4
	M_2	41	6	−1.7070	3	0.8497	2
	M_3	102	4	−1.7026	1	0.8489	1
	M_4	153	2	−1.7042	2	0.8505	3
	M_5	30	7	−1.7513	4	0.9113	6
	M_6	64	5	−1.7536	5	0.9142	7
	M_7	210	1	−1.9629	7	0.8990	5
SZCI	M_1	124	3	−1.9388	6	0.9993	4
	M_2	68	5	−1.8871	3	0.9933	2
	M_3	105	4	−1.8818	1	0.9920	1
	M_4	144	2	−1.8833	2	0.9942	3
	M_5	35	7	−1.9141	4	1.0571	7
	M_6	56	6	−1.9150	5	1.0568	6
	M_7	212	1	−2.0811	7	1.0477	5

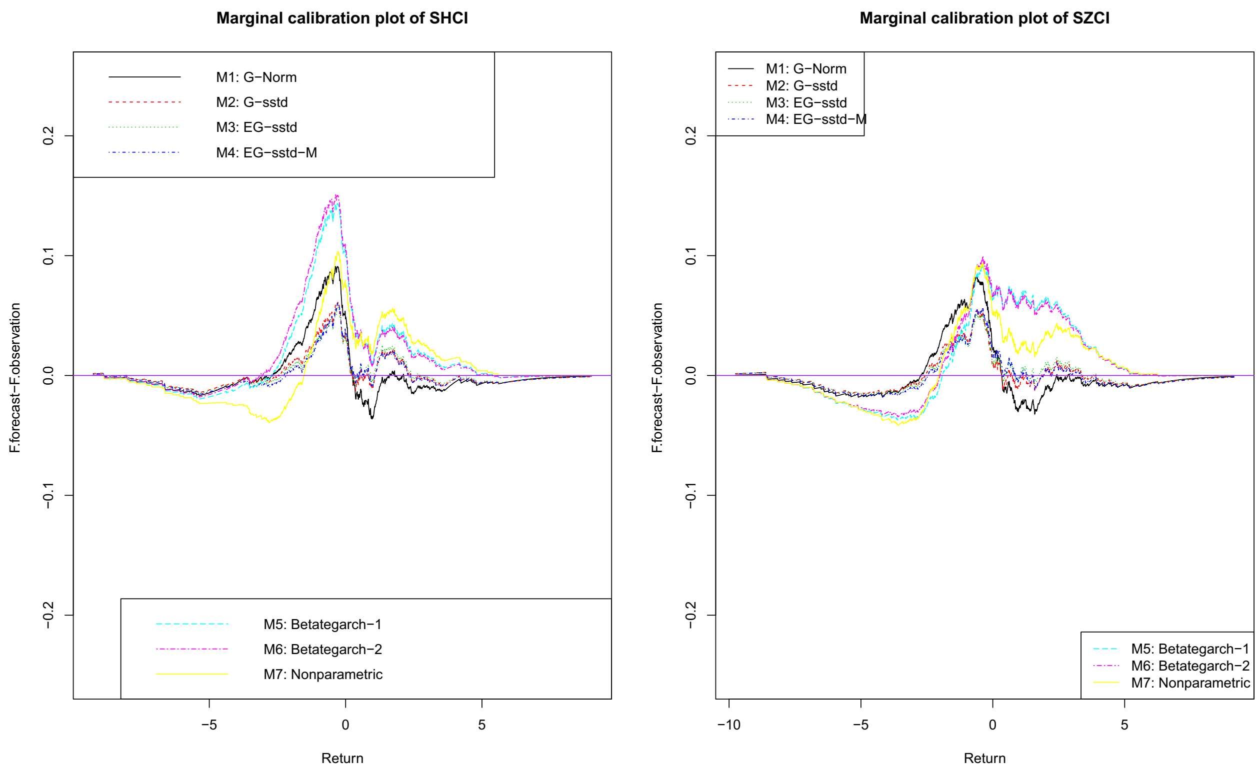

Fig. 1. Marginal calibration charts of SHCI and SZCI.

Table 4. Average width of forecasting intervals.

Index	Interval	M_1	M_2	M_3	M_4	M_5	M_6	M_7
SHCI	50%	2.1591	1.8402	1.8001	1.7964	1.8260	1.8805	1.5542
	90%	5.2647	5.0515	4.9320	4.9332	4.9803	5.1739	4.0838
SZCI	50%	2.5083	2.2171	2.1783	2.1837	1.7078	1.7513	1.9366
	90%	6.1172	5.9295	5.8099	5.8396	4.5539	4.7128	4.9835

the logarithmic likelihood score evaluation, M_7 is the best Bayesian winner, but M_2, M_3, and M_4 have better average logarithmic likelihood scores. M_5, M_6, and M_7 have better sharpness, but M_2, M_3, and M_4 have better average CRPS scores. Due to the different focuses of the models, model uncertainty exists objectively. According to [45], model uncertainty can be reduced using a combination strategy. Hall and Mitchell [46] and Massacci [14] argue that forecasts combination can also resist structural breakpoints. Billio et al. [47] point out that when multiple forecasts can be obtained from different models or sources, these forecasts can be combined to take advantage of all relevant information about the variables to be forecasted, resulting in better forecasting. Pesaran et al. [48] show that model averaging techniques that allow parameter and model uncertainty are particularly important in risk management. For the seven individual models used in this paper, the marginal calibration result of Section 3.2.3 also leads us to consider the combination of models. For the above reasons,three linear dynamic combination strategies are considered, which are equal weight combination (EW), logarithmic score combination (SW) and CRPS combination (CW).

For the forecasting distribution of individual models $F_t^{(j)}(y)$ $(j = 1,2,\ldots,7)$, equal weight combination forecast distribution is defined as

$$F_{com,t}^{EW}(y) = \frac{1}{7}\sum_{j=1}^{7} F_t^{(j)}(y) \quad \ldots\ldots\ldots \quad (26)$$

The logarithmic score combination distribution is:

$$F_{com,t}^{SW}(y) = \sum_{j=1}^{7} \omega_{1t}^{(j)} F_t^{(j)}(y), \quad \ldots\ldots\ldots \quad (27)$$

where $\omega_{1t}^{(j)}$ indicates the weight assigned to the j model at the time t according to likelihood scores, and

$$\omega_{1t}^{(j)} = \frac{\hat{f}_{t|t-1}^{(j)}(y_t)}{\sum_{j=1}^{7} \hat{f}_{t|t-1}^{(j)}(y_t)}, \quad \ldots\ldots\ldots \quad (28)$$

where $\hat{f}_{t|t-1}^{(j)}(y_t)$ is the likelihood score for the j-th model at time t. The dynamic distribution of CRPS weighted is:

$$F_{com,t}^{CW}(y) = \sum_{j=1}^{7} \omega_{2t}^{(j)} F_t^{(j)}(y), \quad \ldots\ldots\ldots \quad (29)$$

where $\omega_{2t}^{(j)}$ indicates the weight assigned to the j-th model at time t, and

$$\omega_{2t}^{(j)} = \frac{\Gamma_{t|t-1}^{(j)}(y_t)}{\sum_{j=1}^{7} \Gamma_{t|t-1}^{(j)}(y_t)}, \Gamma_{t|t-1}^{(j)}(y_t) = \frac{1}{crps(F_{t|t-1}^{(j)}(y_t))}. \quad (30)$$

3.3.2. Direction Predictability and Economic Evaluation

Two-point forecasts are derived from the forecasted distribution of returns, which are mean $\hat{y}_{mean,t}$ and median $\hat{\xi}_{0.5,t}$. The calculation is implemented by the following two formulas:

$$\hat{y}_{mean,t} = \int_{-\infty}^{+\infty} y dF_t(y) \approx \sum_{i=1}^{N-1} y_i[F_t(y_{i+1}) - F_t(y_i)], \quad (31)$$

$$\hat{\xi}_{0.5,t} = inf\{y \in R : F_t(y) \geq 0.5\}. \quad \ldots\ldots\ldots \quad (32)$$

Hence the forecasting accuracy of the direction can be obtained by

$$DA_1 = \frac{1}{T - T_0}\sum_{t=T_0+1}^{T} I(\hat{y}_{mean,t} \cdot y_t > 0), \quad \ldots \quad (33)$$

$$DA_2 = \frac{1}{T - T_0}\sum_{t=T_0+1}^{T} I(\hat{\xi}_{0.5,t} \cdot y_t > 0), \quad \ldots \quad (34)$$

where $T - T_0$ represents the out-of-sample size. Directional predictability is statistically tested by the PT directional accuracy test [20].

As we all know, profit maximization is the goal of return forecasting in financial markets. As pointed by [49] and [14], the forecasting performance of the model should also be evaluated economically through profit-based measures. Referring to [50], the mean transaction return (MTR) of the out-of-sample forecast can be used to evaluate the performance of the trading strategy. The MTR measures the true profit of the financial market when the transaction costs are ignored. According to [50], it is more appropriate to evaluate the performance of the forecasting market movement than the traditional mean square prediction residual or the average absolute forecasting error.

In the case of allowing short selling, if the forecasting of y_t is accurate every time, then the ideal MTR is

$$MTR_{ideal}^{(1)} = \frac{1}{T - T_0}\sum_{t=T_0+1}^{T} |y_t|. \quad \ldots\ldots\ldots \quad (35)$$

Referring to [21], a simple trading strategy is designed, in which a virtual investor can sell short; a point forecast is made based on the distribution of return forecasts: if the forecast for the next period of return is positive, then the trading strategy issues a buy signal; otherwise, it is a sell signal. The MTR of this trading strategy is

$$MTR^{(1)} = \frac{1}{T - T_0}\sum_{t=T_0+1}^{T} \{I(\hat{y}_{t|t-1} > 0) - I(\hat{y}_{t|t-1} \leq 0)\} y_t. \quad (36)$$

The ratio of the strategy MTR to the ideal MTR is defined as

$$Rate^{(1)} = \frac{MTR^{(1)}}{MTR_{ideal}^{(1)}} \times 100\%. \quad \ldots\ldots\ldots \quad (37)$$

The ratio can be used to measure the profitability of trading strategies in the case of short sales. The larger the value, the better the economic benefits. Whether the trading strategy can achieve higher returns can be statistically tested by the AG excess profitability test of [21].

For China's stock market, short selling is not allowed,

Table 5. Direction accuracy and economic evaluation of forecasts (unit: %).

		Mean			Median		
Index	Model	DA_1	$Rate^{(1)}$	$Rate^{(2)}$	DA_1	$Rate^{(1)}$	$Rate^{(2)}$
SHCI	M_1	55.66	4.54	8.69	55.66	4.54	8.69
	M_2	55.66	4.94	9.07	55.66	4.54	8.69
	M_3	54.98	3.26	7.47	55.66	4.54	8.69
	M_4	52.11	−0.05	4.30	54.57	2.93	7.15
	M_5	44.34	−4.54	0	44.34	−4.54	0
	M_6	44.34	−4.54	0	44.34	−4.54	0
	M_7	44.34	−4.54	0	44.34	−4.54	0
	EW	44.75	−1.41	3.00	48.57**	1.08	5.38
	SW	46.79***	6.43***	10.50	64.12***	49.94***	52.12
	CW	46.52***	6.32***	10.39	68.76***	50.39***	52.55
SZCI	M_1	52.93	2.27	4.43	52.93	2.27	4.43
	M_2	52.52	1.28	3.47	52.93	2.27	4.43
	M_3	52.80	1.91	4.08	52.93	2.27	4.43
	M_4	52.80	0.32	2.53	53.21	2.61	7.67
	M_5	47.07	−2.27	0	47.07	−2.27	0
	M_6	47.07	−2.27	0	47.07	−2.27	0
	M_7	47.07	−2.27	0	47.07	−2.27	0
	EW	46.38	−9.44	−7.02	47.75	−6.45	−4.08
	SW	51.57***	15.51***	17.38	58.80***	38.91***	40.26
	CW	49.93***	3.57**	5.71	58.94***	32.84***	34.33

Notes: ** represent significant result at the 5% confidence level.
*** represent significant result at the 1% confidence level.

therefore, the ideal MTR should be amended to:

$$MTR^{(2)}_{ideal} = \frac{1}{T-T_0}\sum_{t=T_0+1}^{T} \max\{y_t, 0\}. \quad . \quad . \quad . \quad . \quad (38)$$

The trading strategy becomes: when the return forecast of the next period is positive, if a unit of asset is owned at hand, then continue to retain it; if there is no asset at hand, then buy one unit of the asset; when the return forecast of the next period is non-positive, if there is a unit of asset at hand, then sell it; if there is no asset at hand, then do nothing and maintain the short position. In fact, the strategy makes the trading always maintains a unit of the asset or a short position. The MTR of the trading strategy is

$$MTR^{(2)} = \frac{1}{T-T_0}\sum_{t=T_0+1}^{T} I(\hat{y}_{t|t-1} > 0) \cdot y_t. \quad . \quad . \quad . \quad (39)$$

The corresponding ratio of the strategy MTR to the ideal MTR is

$$Rate^{(2)} = \frac{MTR^{(2)}}{MTR^{(2)}_{ideal}} \times 100\%. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (40)$$

Table 5 shows the direction accuracy of point forecasts using the mean and median, respectively, of the seven individual models and the three combined models, and the ratios of the strategy MTR to the ideal MTR in both short-selling and not-short-selling environments. The significance of the PT and AG tests is also indicated in the table. In **Table 5**, none of the individual models has significant directional predictability and excess profitability. The EW model is only excessively profitable for SHCI at a significance level of 5%, while all SW and CW models shows significant directional predictability and excess profitability at the 5% level. Furthermore, in addition to the CW model for SZCI under the means forecasting, other models are significant at the 1% level. When the mean is used for point forecasting, the direction accuracy rates of SW and CW are not higher than those of all individual models; surprisingly, the individual models have no significant direction predictability, but SW and CW models are not higher, perhaps because we only perform a simple constant fitting for the conditional mean when modeling, and when the rolling window is too wide, it shows that the "one-sided" is all positive or negative, so the result of the test is not significant direction predictability, while the combined model is not. In the case where the direction accuracy rate is lower than the individual model, the SW and CW models can also have a higher excess profit than the individual model, which may be because the two combined models have a more favorable distribution of return in the case of a sharp rise and fall. When using the median for point forecasting, we can find that the direction accuracies of SW and CW models are significantly higher than that of any individual model. Whether it is short-selling or not, the ratio of the strategy MTR to the ideal MTR far exceeds that of any individual model. The combined models of SW and CW have significant directional accuracy and higher excess profitability than the individual models, which means that the appropriate model combination may be closer to the real data generation process, and from the perspective of economic evaluation, they may have better forecasting performance for the conditional distribution of returns.

4. Conclusions

In this paper, the predictability of stock market returns is studied from the perspective of distribution forecasting. Seven individual models are used, namely, GARCH-norm, GARCH-sstd, EGARCH-sstd, EGARCH-sstd-M, one-component Beta-t-EGARCH, two-component Beta-t-EGARCH, and EWMA-based nonparametric models. The rolling window modeling and out-of-sample forecasting are performed on the conditional distribution of the returns of SHCI and SZCI. PIT evaluation, the average log-likelihood scores, and the average CRPS scores show that GARCH-sstd, EGARCH-sstd, and EGARCH-sstd-M are superior to the other models. However, the nonparametric model is the best Bayesian winner, and the one-component Beta-t-EGARCH, the two-component Beta-t-EGARCH and the nonparametric model have better tail calibration and sharpness. However, no model can pass the statistical test completely, that is, no one is an absolute qualified distribution forecasting model. Furthermore, none of the individual models has both predictability power and excess profitability.

Three combination strategies of the models are designed, which are equal weight combination (EW), log-likelihood score combination (SW), and CRPS score combination (CW). It was found that SW and CW models have significant direction predictability and excess profitability compared to the individual model, whether it is for SHCI or SZCI. It indicates that the log-likelihood score combination and the CRPS score combination may be closer to the real data generation process, and from the perspective of economic evaluation, the SW and CW models maybe have a predictive effect on the conditional return distribution in China's stock market.

Acknowledgements

This work is supported by the Philosophy and Social Science Planning Project of Zhejiang Province (20NDJC142YB), the Humanities and Social Sciences Research Project of Ministry of Education of China (15YJC630192), the National Social Science Fund of China (15CGL009), the Hangzhou Social Science Planning Project for Talent Cultivation (2017RCZX25), and the 2020-2021 Research Project of Shaoxing University (Research on Forecasting Method and Application of Financial Return Distribution Based on Model Selection).

References:

[1] E. F. Fama, "Random walks in stock market prices," Financial Analysts J., No.21, pp. 55-59, 1965.

[2] A. W. Lo, "The adaptive markets hypothesis," The J. of Portfolio Management, No.5, pp. 15-29, 2004.

[3] E. F. Fama, "Efficient Capital Markets: II," The J. of Finance, Vol.46, No.5, pp. 1575-1617, 1991.

[4] J. H. Cochrane, "New facts in finance," Economic Perspectives, Vol.23, No.3, pp. 36-58, 1999.

[5] K. P. Lim, W. Luo, and J. H. Kim, "Are US stock index returns predictable? evidence from automatic autocorrelation-based tests," Applied Economics, Vol.45, No.8, pp. 953-962, 2013.

[6] X. Zhou, Q. Song, and X. Liu, "Adaptive markets hypothesis and evidence from China's capital markets," J. of Management Science, No.6, pp. 111-126, 2017 (in Chinese).

[7] J. H. Cochrane, "The dog that did not bark: a defense of return predictability," Review of Financial Studies, Vol.21, No.4, pp. 1533-1575, 2008.

[8] W. E. Ferson, S. Sarkissian, and T. Simin, "Asset pricing models with conditional betas and alphas: the effects of data snooping and spurious regression," J. of Financial and Quantitative Analysis, Vol.43, No.2, pp. 331-353, 2008.

[9] J. T. Campbell and P. Samuel, "Predicting excess stock returns out of sample: can anything beat the historical average?," Scholarly Articles, Vol.21, No.4, pp. 1509-1531, 2008.

[10] Y. Zhang, Q. Zeng, F. Ma, and B. Shi, "Forecasting stock returns: do less powerful predictors help?," Econ. Modell., doi: 10.1016/j.econmod.2018.09.014, 2018.

[11] T. Dangl and M. Halling, "Predictive regressions with time-varying coefficients," J. Financ. Econ., Vol.106, No.1, pp. 157-181, 2012.

[12] X. Zhu and J. Zhu, "Predicting stock returns: a regime-switching combination approach and economic links," J. Bank. Finance, Vol.37, No.11, pp. 4120-4133, 2013.

[13] D. E. Rapach, J. K. Strauss, and G. Zhou, "Out-of-sample equity premium prediction: combination forecasts and links to the real economy," Rev. Financ. Stud., Vol.23, No.2, pp. 821-862, 2010.

[14] D. Massacci, "Predicting the distribution of stock returns: model formulation, statistical evaluation, VaR analysis and economic significance," J. of Forecasting, No.34, pp. 191-208, 2015.

[15] J. Zheng, H. Pan, and Y. Fan, "Analysis of return directional predictability of Chinese stock market based on generalized cross-spectrum approach," China Soft Science, No.8, pp. 154-160, 2010 (in Chinese).

[16] P. F. Christoffersen and F. X. Diebold, "Financial Asset Returns, Direction-of-Change Forecasting, and Volatility Dynamics," Management Science, Vol,52, pp. 1273-1287, 2006.

[17] L. Li, S. Guan, and M. Guan, "Economic analysis of time-varying distribution of stock market returns and empirical evidence in China," Statistical Research, No.11, pp. 66-78, 2011 (in Chinese).

[18] Y. Yao and B. Xu, "Conditional distribution prediction of stock returns and its application on risk aversion analysis," J. Adv. Comput. Intell. Intell. Inform., Vol. 22, No.4, pp. 448-456, 2018.

[19] L. Liu, R. Bu, Z. Pan, and Y. Xu, "Are financial returns really predictable out-of-sample? Evidence from a new bootstrap test," Economic Modelling, Vol.81, pp. 124-135, 2019.

[20] M. H. Pesaran and A. Timmermann, "A simple nonparametric test of predictive performance," J. of Business and Economic Statistics, Vol.10, No.4, pp. 461-465, 1992.

[21] S. Anatolyev and A. Gerko, "A trading approach to testing for predictability," J. of Business and Economic Statistics, Vol.23, No.4, pp. 455-461, 2005.

[22] M. H. Pesaran and A. Timmermann, "Real-time econometrics," Econometric Theory, Vol.21, No.1, pp. 212-231, 2005.

[23] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," Econometrica, Vol.50, No.4, pp. 987-1007, 1982.

[24] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," J. of Econometrics, No.31, pp. 307-327, 1986.

[25] P. Wang and J. Wang, "The return distribution of China's stock market and its SPA test," J. of Systems Management, Vol.17, No.5, pp. 542-547, 2008 (in Chinese).

[26] D. B. Nelson, "Conditional heteroskedasticity in asset returns: A new approach," Econometrica, Vol.59, No.2, pp. 347-370, 1991.

[27] R. F. Engle, D. M. Lilien, and R. P. Robins, "Estimating time varying risk premia in the term structure: The arch-m model," Econometrica: J. of the Econometric Society, Vol.55, No.2, pp. 391-407, 1987.

[28] A. C. Harvey and T. Chakravarty, "Beta-t-(E)GARCH," Cambridge Working Papers in Economics 0840, Faculty of Economics, University of Cambridge, 2008.

[29] A. C. Harvey, "Dynamic models for volatility and heavy tails," Cambridge University Press, 2013.

[30] A. C. Harvey and G. Sucarrat, "EGARCH models with fat tails, skewness and leverage," Computational Statistics and Data Analysis, No.76, pp. 320-338, 2014.

[31] S. Blazsek and M. Villatoro, "Is Beta-t-EGARCH(1,1) superior to GARCH(1,1)," Applied Economics, Vol.47, No.17, pp. 1764-1774, 2015.

[32] S. Blazsek and V. Mendoza, "QARMA-Beta-t-EGARCH versus ARMA-GARCH: an application to S&P 500," Applied Economics, Vol.48, No.12, pp. 1119-1129, 2016.

[33] Z. Ding, R. Engle, and C. Granger, "A long memory property of stock market returns and a new model," J. of Empirical Finance, No.1, pp. 83-106, 1993.

[34] A. Harvey and V. Oryshchenko, "Kernel density estimation for time series data," Int. J. of Forecasting, No.28, pp. 3-14, 2012.

[35] G. Elliott and A. Timmermann, "Economic forecasting," Princeton University Press, 2016.

[36] A. Pérez, "Comments on "Kernel density estimation for time series data"," Int. J. of Forecasting, No.28, pp. 15-19, 2012.

[37] J. Nowotarski and R. Weron, "Recent advances in electricity price forecasting: A review of probabilistic forecasting," Renewable and Sustainable Energy Reviews, No.81, pp. 1548-1568, 2018.

[38] A. P. Dawid, "Statistical theory: the prequential approach," J R Stat Soc A, Vol.147, No.2, pp. 278-292, 1984.

[39] F. X. Diebold, A. G. Todd, and S. T. Anthony, "Evaluating density forecasts with applications to financial risk management," Int. Economic Review, Vol.39, No.4, pp. 863-883, 1998.

[40] J. Berkowitz, "Testing density forecasts, with applications to risk management," J. of Business and Economic Statistics, Vol.19, No.4, pp. 465-474, 2001.

[41] A. Ghalanos, "Introduction to the rugarch package (Version 1.3-8)," 2020, https://cran.r-project.org/web/packages/rugarch/vignettes/Introduction_to_the_rugarch_package.pdf [accessed February 7, 2020]

[42] Y. Hong and H. Li, "Nonparametric specification testing for continuous-time models with applications to term structure of interest rates," Review of Financial Studies, Vol.18, No.1, pp. 37-84, 2005.

[43] Y. Ait-Sahalia, "Testing continuous-time models of the spot interest rate," Review of Financial Studies, Vol.9, No.2, pp. 385-426, 1996.

[44] T. Gneiting, F. Balabdaoui, and A. E. Raftery, "Probabilistic forecasts, calibration and sharpness," J. R. Statist. Soc. B, No.69, pp. 243-268, 2007.

[45] A. Timmermann, "Forecast combinations," G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), "Handbook of Economic Forecasting," Volume 1, pp. 135-196, Elsevier, 2006.

[46] S. G. Hall and J. Mitchell, "Combining density forecasts," Int. J. of Forecasting, Vol.23, No.1, pp. 1-13, 2007.

[47] M. Billio, R. Casarin, F. Ravazzolo, and K. D. Herman, "Time-varying combinations of predictive densities using nonlinear filtering," J. of Econometrics, Vol.177, Issue 2, pp. 213-232, 2013.

[48] M. H. Pesaran, C. Schleicher, and P. Zaffaroni, "Model averaging in risk management with an application to futures markets," J. of Empirical Finance, Vol.16, No.2, pp. 280-305, 2009.

[49] G. Leicht and J. E. Tanner, "Economic forecast evaluation: profits versus the conventional error measures," American Economic Review, Vol.81, No.3, pp. 580-590, 1991.

[50] J. Fan and Q. Yao, "Nonlinear time series: modeling, forecasting and application," Higher Education Press, 2005 (in Chinese).

Name:
Yanyun Yao

Affiliation:
Lecturer, Department of Applied Statistics, Shaoxing University

Address:
900 Chengnan Avenue, Yuecheng District, Shaoxing 312000, China

Brief Biographical History:
1998-2002 Bachelor Course of Applied Mathematics, Beijing Institute of Technology
2002-2005 Master Course of Operational Research and Cybernetics, Beijing Institute of Technology
2012-2019 Doctor Course of Quantitative Economics, Zhejiang Gongshang University
2005- Lecturer, Department of Mathematics, Shaoxing University

Main Works:
• Y. Yao and B. Xu, "Conditional Distribution Prediction of Stock Returns and its Application on Risk Aversion Analysis," J. Adv. Comp. Intell. Intell. Inform., Vol.22, No.4, pp. 448-456, 2018.

Name:
Xiutian Zheng

Affiliation:
Associate Professor, Qianjiang College, Hangzhou Normal University

Address:
16 Xuelin Street, Xiasha Education Park, Hangzhou, Zhejiang 310018, China

Brief Biographical History:
2002-2006 Bachelor Course, Finance, Zhejiang Gongshang University
2006-2009 Master Course, Quantitative Economics, Zhejiang Gongshang University
2010-2016 Doctor Course, Enterprise Management, Zhejiang Gongshang University
2016-Associate Professor, Qianjiang College, Hangzhou Normal University
2020-Postdoctoral fellow, School of Management, Fudan University

Main Works:
• X. Zheng, Y. Xu, and L. Gu, "Impact of Venture Capital Firms' Reputation on Investment Decisions," J. of Applied Sciences,Vol.13, No.14, pp.2694-2699, 2013.

Name:
Huimin Wang

Affiliation:
Associate Professor, Department of Applied Statistics, Shaoxing University

Address:
900 Chengnan Avenue, Yuecheng District, Shaoxing 312000, China

Brief Biographical History:
1998-2002 Bachelor Course of Applied Mathematics, Northeastern University
2002-2005 Master Course of Applied Mathematics, Zhejiang University
2009-2012 Doctor Course of Applied Mathematics, Zhejiang University
2005- Department of Mathematics, Shaoxing University

Main Works:
• H. Wang and S. Li, "Stable Recovery of Low Rank Matrices From Nuclear Norm Minimization," Acta Mathematicae Applicatae Sinica, English Series, Vol.31, No.1, pp. 1-14, 2015.