Economic Research-Ekonomska Istraživanja

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rero20

# Daily return distribution forecast incorporating intraday high frequency information in China's stock market

Yanyun Yao, Qiang Huang & Shangzhen Cai

**To cite this article:** Yanyun Yao, Qiang Huang & Shangzhen Cai (2023) Daily return distribution forecast incorporating intraday high frequency information in China's stock market, Economic Research-Ekonomska Istraživanja, 36:2, 2107554, DOI: 10.1080/1331677X.2022.2107554

**To link to this article:** https://doi.org/10.1080/1331677X.2022.2107554



Published online: 16 Aug 2022.

ECONOMIC RESEARCH-EKONOMSKA ISTRAŽIVANJA
2023, VOL. 36, NO. 2, 2107554
https://doi.org/10.1080/1331677X.2022.2107554

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

# Daily return distribution forecast incorporating intraday high frequency information in China's stock market

Yanyun Yao[a,b], Qiang Huang[b] and Shangzhen Cai[c]

[a]College of Business Administration, Ningbo University of Finance & Economics, Ningbo, China; [b]Department of Mathematics, Shaoxing University, Shaoxing, China; [c]College of Digital Technology and Engineering, Ningbo University of Finance & Economics, Ningbo, China

**ABSTRACT**

The stock market forecast is an important and challenging issue. Its distribution forecast of returns can provide information that is more complete, compared to point forecast and interval forecast. As intraday high-frequency information is available, we incorporate intraday returns into the predictive modelling of daily return distribution in two ways: realized volatility and scale calibration. Three parametric models, EGARCH, EGARCH-X, and realGARCH, and two nonparametric models, NP and realNP, are used. Our improved NP model, the realNP model, is based on intraday returns calibration. The results show that intraday information improves goodness-of-fit and forecasting effect, and the realGARCH model is relatively the best. According to the realNP model results, the intraday returns can only contribute about a 30% description of the daily distribution and less than 1% information for a one-step-ahead distribution forecast. Furthermore, three combinations are considered, and the log-score and CRPS combinations are found to have direction predictability and excess profitability statistically. The non-short-selling situation consistently has more excess profits than the short-selling situation, which implies that the non-short-selling rule protects investors. This study reveals the importance of incorporating intraday information and model combinations for stock market forecast modelling.

**ARTICLE HISTORY**
Received 6 January 2022
Accepted 26 July 2022

**KEYWORDS**
Intraday information; distribution forecast; realized volatility; scale calibration; model combination

**JEL CODES**
C51; C53; G17

## 1. Introduction

The stock market is an important embodiment of a country's economic and financial activities. It plays an important role in the economic development of various countries and has an important impact on countries, enterprises, and individuals. If the stock market evolution process is modelled properly, it will better ensure efficient resource allocation. Investment institutions and investors benefit from early detection of stock market movements. However, the stock market is a typical complex system

**CONTACT** Shangzhen Cai 59763489@qq.com

with great uncertainty and volatility (Hwan et al., 2011). Therefore, stock market forecast is an important and challenging issue.

Financial variables, such as stock prices, are typically available daily, and, in the past decade, the availability of intraday level (high frequency) information has increased dramatically. Therefore, research on high frequency data is extensive. High-frequency data allow the study on intraday time scales and may contain valuable information for longer time scales, which may be more meaningful to most market participants. Consequently, many studies have attempted to incorporate intraday data into the modelling and forecasting of daily and even lower frequency variables. Naturally, financial market behaviour is investigated using mixed-sampling data. For example, Gorgi et al. (2019) constructed a flexible and easy-to-implement framework to forecast low-frequency time series variables using timely information from high-frequency variables.

Regarding the application of high-frequency data, a large category involves the realized measures, including the traditional realized volatility measures and the generalized realized measures (Jiang et al., 2018; Jiang & Gu, 2019). The traditional realized volatility measures include Realized Variance (RV) (Andersen & Bollerslev, 1998), micro-noise robust Realized Kernel (Barndorff-Nielsen et al., 2008), and jump-robust Realized Bipower Variation (Barndorff-Nielsen & Shephard, 2004). Generalized realized measures include Realized Absolute Dispersion (Tsay, 2010) and realized Lower Partial Moment (Huffman & Moll, 2012), as well as realized value at risk (VaR) and realized expected shortfall (ES) (Wuertz et al., 2009). RV, introduced by Andersen and Bollerslev (1998) and formally determined by Andersen et al. (2001), is the most basic and commonly used measure. Compared with the model that only uses daily data, the use of realized volatility measurement has been found to improve the prediction of daily return volatility significantly (Wang & Huang, 2015; Lyócsa et al., 2021; Lu et al., 2021). Volatility involves only the first two moments of returns. However, many financial situations have insufficient information on the first two moments of returns. For example, in risk management, to calculate the VaR or the ES, it is necessary to understand the specific quantile of the return distribution. In addition, numerous studies have shown that higher-order moments, such as skewness and kurtosis are time varying, and are related to portfolio allocation and asset pricing (Brooks et al., 2005). Therefore, it is necessary to extend the return modelling that incorporates high-frequency information to a higher-order moment or distribution prediction level.

English documents such as Andersen et al. (2003), Clements et al. (2008), Maheu and McCurdy (2011), and Hansen et al. (2012), as well as Chinese documents such as Wang and Huang (2012a, 2015), Huang et al. (2015), Zhu and Tang (2017), and Yu and Wang (2018), extend the use of volatility measures to the quantile or distribution level. In such studies, the realized volatility is linked with the density (or quantile) of daily returns. It includes two components: First, a parametric time series modelling of volatility, which includes some realized volatility measurements, to simulate and generate a daily volatility point forecast. The second component is the assumption of the parameter distribution of daily returns, which allows the density or quantile forecast of daily returns to be generated from the point forecast of daily-realized volatility. Commonly used models are GARCH (Wang & Huang, 2015), MIDAS (Clements et al., 2008), and HAR (Maheu & McCurdy, 2011; Zhu & Tang, 2017), among others. Owing to the excellent performance of the GARCH model in traditional volatility modelling, the combination of realized volatility and the GARCH model has become a major research focus.

Regarding the limitations of the above methods, Hallam and Olmo (2014a) proposed a density prediction method for a single fractal. Under the assumption of a single fractal, the form of the distribution scale can be estimated using a given data sample, and these estimates can be used to accurately rescale intraday returns so that they are equal under the daily return distribution; then, the density of daily returns can be directly estimated from these rescaled intraday observations. However, the empirical application of Hallam and Olmo (2014a) shows that with large assumption deviation from the single fractal distribution, the predictive ability of the single fractal method may be limited. Furthermore, Hallam and Olmo (2014b) proposed a more general and flexible multi-fractal density prediction method. This method overcomes the key theoretical limitations of the single fractal method and allows for scaling relationships between the return distributions to be more flexible under different sampling frequencies. The variation range and sign of the intraday returns can be incorporated into the estimation of the daily return density, allowing the intraday data to directly affect moments that are higher than the second moment of the daily returns. However, the implementation of this method has limitations, that is, it needs to set a specific parameter form for the daily return distribution.

This study aims to model the conditional distribution forecast of daily returns. We hope to take advantage of the GARCH model and consider the parametric method that combines intraday high-frequency information with the GARCH model. The EGARCH model proposed by Nelson (1991), which is widely used in the literature, is used as the benchmark parametric model. The GARCH-X model adds the realized volatility, as an exogenous variable, into the conditional variance equation, and the realGARCH model, which combines the realized volatility and the daily-implied volatility, is included. However, to reduce the risk of model 'misspecification', nonparametric models are considered. The model proposed by Harvey and Oryshchenko (2012) (short form, NP) serves as a benchmark for nonparametric models. Furthermore, we consider intraday returns calibration and integrate them into NP through a recursive mechanism. To reflect the impact of intraday high-frequency information, model comparison is adopted. Model combinations are also considered to reduce the bias of individual models. Statistical evaluation and economic evaluation are both conducted in this study.

In general, this study has three innovations. (1) Based on the scale calibration, a new model, realNP, is proposed, which can effectively quantify the influence of intraday information on daily return distribution and one-step-ahead distribution forecast. (2) The importance of intraday high-frequency information is considered from the perspective of overall distribution forecast through model comparison. The distribution forecast is more complete, with credible results. (3) Statistical evaluation results of the distribution forecast are used to construct combinations of parametric and non-parametric models, and to find valuable combined models, directional predictability and excess profitability are considered as economic evaluation.

The remainder of this paper is organized as follows. Section 2 presents the data description and information pre-processing. Section 3 introduces the models and methods. Section 4 presents the empirical results, including the full-sample parameter estimation and fitting results, statistical evaluation of the rolling-sample distribution forecast, and model combinations and economic evaluation. In Section 5, a robustness test is conducted. Finally, Section 6 presents the conclusions, some limitations, and scope for further research.

## 2. Data description and information pre-processing

### 2.1. Daily returns and intraday high frequency returns

The Shanghai Composite Index (SHCI) is often used to represent China's stock market (Yao et al., 2021), and it was selected as the research object. Since 2010, China's stock market has launched margin and securities lending businesses and stock index futures trading. Therefore, profits are not only made unilaterally, as per the history of China's stock market. The evolution of return distribution after 2010 may be different from that before 2010. Therefore, this empirical study was conducted from 4 January 2010 to 29 December 2017. The return is calculated as the closing price's logarithmic return percentage, namely,

$$y_t = (\ln p_t - \ln p_{t-1}) \times 100, t = 1, 2, \ldots, T, \tag{1}$$

where $p_t$ represents the closing price of SHCI on day t, and $p_0$ corresponds to the closing price on the trading day before 4 January 2010. The sample size of the returns is $T = 1940$. To realize the return distribution forecast, the rolling samples are adopted. The initial fitting sample comprises data from 4 January 2010 to 31 December 2014. Therefore, the in-sample size is $T_0 = 1208$. The out-of-sample forecast period is from 5 January 2015 to 29 December 2017: the out-of-sample size is $n = T - T_0 = 732$. The robustness test uses the returns from 4 January 2010 to 30 June 2021, and $T = 2788$. The initial fitting sample comprises data from 4 January 2010 to 29 December 2017 (eight years), that is $T_0 = 1940$. 2 January 2018 to 31 June 2021 is the out-of-sample prediction interval, that is $n = T - T_0 = 848$. Figure 1 shows the closing price data for the Shanghai Composite Index from 4 January 2010 to 30 June 2021. The period between the solid red lines is the empirical forecast time interval. This covers three market states: 'consolidation', 'bull', and 'bear'. The blue dotted lines are the forecast periods for the robustness test, including two market states of 'consolidation' and 'bear'.

We hope to utilize the potential value of the intraday trading information fully, and carefully avoid the interference of market microstructure noise caused by overly short sampling intervals. Therefore, the sampling frequency problem is worth considering, and Hansen and Lunde (2006), Bandi and Russell (2008), and Wang and Nishiyama (2015) have discussed it. This problem has also been encountered in the studies on realized volatility, such as Andersen et al. (2001), which find that the 5-minute sampling interval is a good trade-off between the above two factors. Many documents on intraday high-frequency information, such as Maheu and McCurdy


**Figure 1.** The SHCI's daily closing price and forecast period.
*Source:* compiled by authors with the help of R software.

(2011) and Hallam and Olmo (2014a, 2014b), also use 5-minute trading data. Therefore, this study selects intraday 5-minute trading data. The time span for intraday returns is similar to that for the daily return: from 4 January 2010 to 29 December 2017. There are $N_t$ 5-minute closing prices on the t-th trading day, and similar to the daily return, the intraday return is defined by a 'closing price-closing price' form, that is

$$r_{t,i} = (\ln p_{t,i} - \ln p_{t,i-1}) \times 100, t = 1, 2, \ldots, T, i = 1, 2, \ldots, N_t, \tag{2}$$

where $r_{t,i}$ and $p_{t,i}$ represent the i-th 5-minute return and closing price, respectively, on the t-th day. In particular, the first 5-minute closing price of each day is compared with the last 5-minute closing price of the previous day, that is, $p_{t,0} = p_{t-1,N_{t-1}}$; therefore, this intraday return is also called the overnight return. It can be seen that $N_t$ also represents the sample size of high-frequency returns on day t.

As China's stock market trading time is 9:30-11:30 and 13:00-15:00, $N_t = 48$, and there are N $= N_t \times T =$ 931205-minute high-frequency returns. What are the characteristics of intraday high-frequency returns? Andersen and Bollerslev (1997) and Engle and Sokalska (2012) proposed that the intraday returns of regular sampling are 'seasonal'. For a deeper understanding of intraday returns, we average the returns at the same time point, from 9:35 to 15:00, a total of 48 5-minute average returns. As shown in Figure 2, the returns are abnormal when the market opens, closes in the morning, and closes in the afternoon. Compared with the previous day's closing, the opening of the market shows a relatively 'pessimistic' performance, and the return is significantly negative, while the closing of the market in the morning and afternoon shows a state of 'excitement', with significant positive returns; however, for other time points, most of these fluctuate around 'zero', with no obvious positive or negative characteristic.

## *2.2. Realized volatility*

Realized volatility is the most commonly used to model intraday high-frequency returns. The simplest realized volatility measurement is proposed by Andersen and Bollerslev (1998) and is defined as

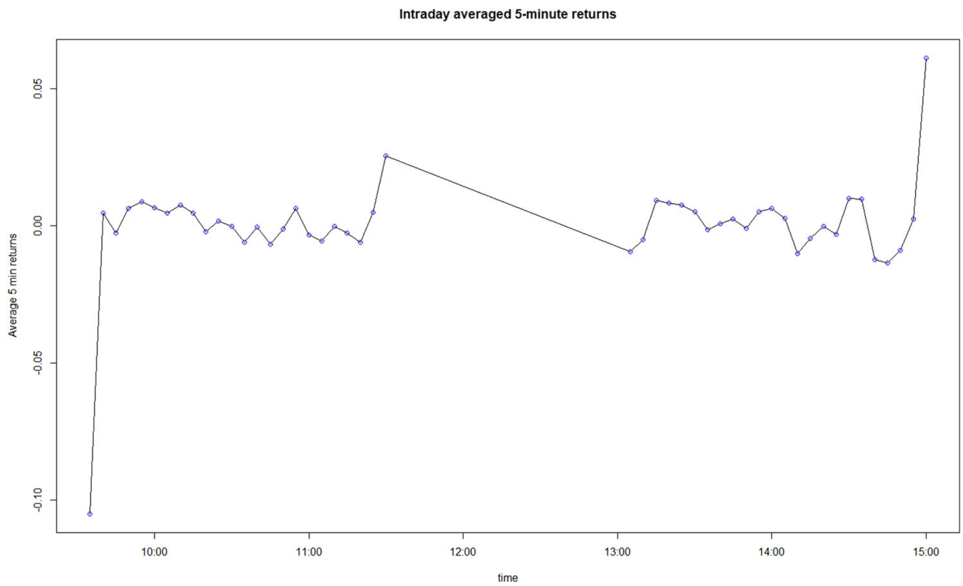


**Figure 2.** The intraday averaged 5-minute returns.
*Source:* compiled by authors with the help of R software

$$RV_t = \sum_{i=1}^{N_t} r_{t,i}^2. \quad (3)$$

It uses $RV_t$ and its logarithm $\ln(RV_t)$ for model construction. Figure 2 depicts the square of daily returns and the daily-realized volatility, and shows that the realized volatility describes the fluctuation characteristics of the square of return. (Figure 3)

## 2.3. Scale calibration of intraday high frequency returns

Intraday high frequency information modelling is used because of the conjecture that 'daily returns and intraday high-frequency returns may have a certain relationship'. What is the relationship between daily return distribution and intraday high frequency return distribution? First, descriptive statistics is performed on daily return $\{y_t\}$ and intraday return $\{r_{t,i}\}$. Based on the maximum and minimum values shown in Table 1, the two seem 'equal', but from the 25% and 75% quantiles, the difference between them is very big. The interquartile range of daily return (75% quantile–25% quantile) is 1.2018, while the interquartile range of 5-minute return is only 0.1507. The differences between their median and mean are also great. Further, as shown in Figure 4, the difference between the kernel densities of daily return $\{y_t\}$ and 5-minute return $\{r_{t,i}\}$ is obvious. The daily return is 'shorter and fatter', the expansion range is wider, and the maximum does not exceed 0.5, while the 5-minute returns are mostly concentrated between $-1$ and 1, with the maximum being close to 4. This scale difference is mentioned by Hallam and Olmo (2014a, 2014b). To establish the relationship between the daily return and the 5-minute return, the 5-minute return

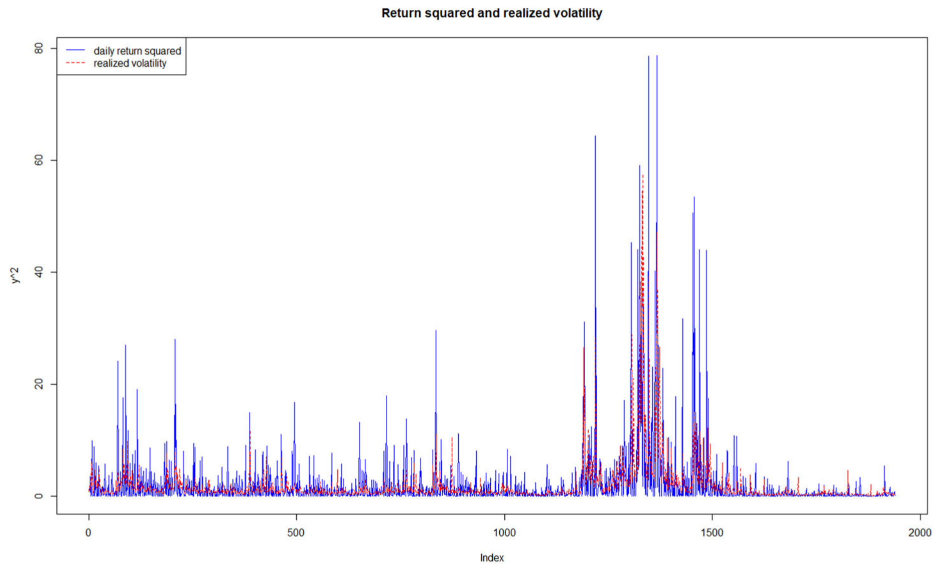


**Figure 3.** The square of daily returns and realized volatility.
*Source:* compiled by authors with the help of R software

**Table 1.** Descriptive statistics of daily return and 5-minute return.

| Type | Minimum | 25% quantile | Median | Mean | 75%quantile | Maximum |
|---|---|---|---|---|---|---|
| Daily return | −8.8729 | −0.5650 | 0.0595 | 0.0026 | 0.6368 | 5.6036 |
| 5-min return | −6.9193 | −0.0735 | 0.0012 | 0.00001 | 0.0772 | 5.7431 |

Source: calculated by authors via R software.

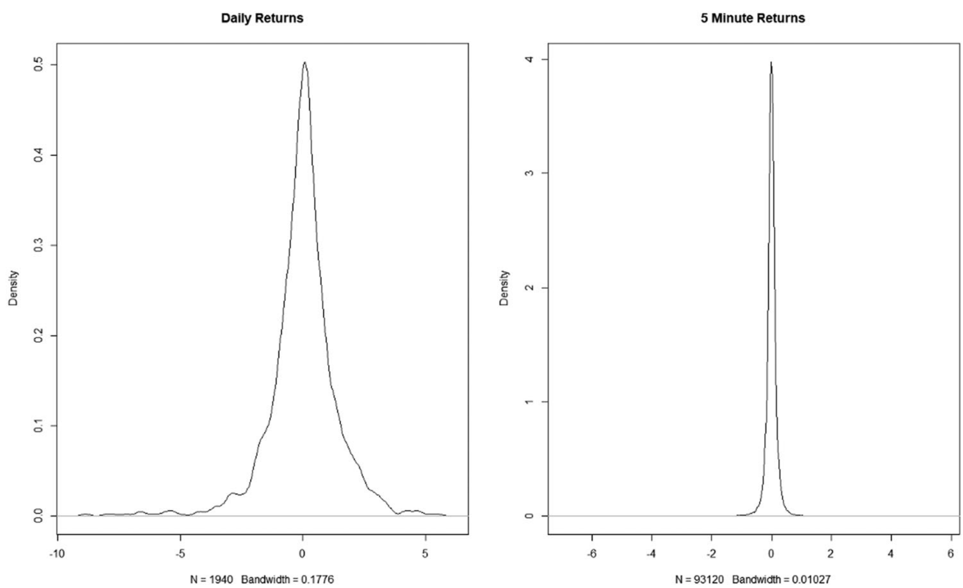


**Figure 4.** The kernel density of daily return and 5-minute return.
*Source:* compiled by authors with the help of R software

should be rescaled so both returns have the same 'scale'. Hallam and Olmo (2014a, 2014b) proposed rescaling methods for intraday returns for single-fractal and multi-fractal scenarios; however, proving whether intraday returns follow single-fractal or multi-fractal distribution is difficult. Therefore, this paper proposes different modelling ideas based on the central limit theorem.

As our return model adopts the logarithmic 'closing price - closing price' model, the relationship between daily return and the intraday return is as follows:

$$y_t = \sum_{i=1}^{N_t} r_{t,i}. \tag{4}$$

From Figure 2, except for the three time-points of opening and closing in the morning, and closing in the afternoon, the intraday returns at other time-points can be regarded as having the same 'fluctuation pattern'. Therefore, the intraday returns are assumed as independent and identically distributed (iid). According to the central limit theorem, when $N_t$ is large enough, then

$$y_t \sim N\left(N_t \cdot E(r_{t,i}),\ N_t \cdot Var(r_{t,i})\right), \tag{5}$$

where $E(r_{t,i})$ and $Var(r_{t,i})$ represent the mean and variance of the intraday returns, and $\mathrm{N}(\cdot)$ indicates normal distribution. As such, we only need to estimate the mean and variance of intraday returns to construct the distribution of the daily returns. $E(r_{t,i})$ can be estimated by sample mean, and $Var(r_{t,i})$ can be estimated by sample variance, namely,

$$\mathrm{E}(r_{t,i}) \approx \mu_{t,i} = \frac{1}{N_t}\sum_{i=1}^{\mathrm{N}_t} r_{t,i}, \tag{6}$$

$$\mathrm{Var}(r_{t,i}) \approx S_{t,i}^2 = \frac{1}{N_t - 1}\sum_{i=1}^{N_t} (r_{t,i} - \mu_{t,i})^2. \tag{7}$$

Here, $N_t$ is expected to be larger, but a larger $N_t$ indicates a shortened sampling interval, which will be interfered with by microstructure noise. The use of a 5-minute return means $N_t = 48$. $N_t = 48$ is considered 'large enough'. The value of $N_t$ and the iid assumption of the intraday returns may have some potential deviations. Therefore, the above-mentioned normal distribution pattern of daily return $y_t$ may be regarded as 'information' in the model construction, rather than the whole distribution of $y_t$.

## 3. Models and methods

To investigate the impact of intraday high-frequency information on the distribution of daily returns, parametric and nonparametric methods are both studied, based on model comparison. There are three parametric models. The first is the EGARCH model, which performs well, based on the literature. The second is the EGARCH-X model, which directly adds the realized volatility to the variance equation. The third is the realGARCH model proposed by Hansen et al. (2012) and improved by Wang and Huang (2012a). According to the literature, in the parametric models, the mean equation is modelled through ARMA processing, and the residual distribution is set as a skewed student's t distribution (sstd) (Jiang & Gu, 2019).

The nonparametric model comprises two models: NP and realNP. The NP model, proposed by Harvey and Oryshchenko (2012), models the entire return distribution rather than the first or second-order moments. The realNP model is this study's innovative model, which integrates intraday high frequency returns into the NP model through scale calibration.

## 3.1. Parametric models

### 3.1.1. The EGARCH model

The EGARCH model allows leverage effects, removes non-negative constraints of the parameters to be estimated, and is more flexible in estimation. Many documents have shown that it has a good fitting effect on security returns. Owing to the above advantages, increasing security empirical studies adopt it to perform modelling. The EGARCH model with sstd residuals often performs relatively better. Therefore, it is selected as the parametric benchmark model in this paper. As ARMA modelling of the conditional mean may have certain benefits, the ARMA(1,1) is used to model the conditional mean, and the overall model is expressed as follows:

$$\begin{cases} y_t = \mu + \delta y_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t, \\ \ln h_t = \omega + \beta \ln h_{t-1} + \alpha \left| \dfrac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} \right| + \gamma \dfrac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}}, \\ \varepsilon_t | F_{t-1} \overset{iid}{\sim} \dfrac{2}{(\lambda + \lambda^{-1})\sqrt{h_t}} \left[ f_v\left( \dfrac{\varepsilon_t}{\lambda\sqrt{h_t}} \right) \cdot I(\varepsilon_t \geq 0) + f_v\left( \dfrac{\varepsilon_t}{\lambda\sqrt{h_t}} \right) \cdot I(\varepsilon_t < 0) \right]. \end{cases} \tag{8}$$

In the EGARCH model, $y_t$ represents the daily return, $h_t$ is the conditional heteroscedasticity of $y_t$, and $\gamma$ is the leverage effect parameter. $F_{t-1}$ represents the information set at time t-1, $f_v(\cdot)$ represents the density function of the student t distribution with the freedom degree $\nu$, $\lambda$ is the skewness coefficient, and I$(\cdot)$ represents the indicative function.

### 3.1.2. The EGARCH-X model

High-frequency data has become increasingly accessible, which has led to many 'realization' metrics based on high-frequency data, and the realized volatility in Section 2.2 is the traditional and most commonly used one. How is the realized volatility integrated into the modelling of return distribution? As EGARCH has good performance, the lagging realized volatility is included naturally as an 'exogenous variable' to the conditional variance equation. The earliest research on this method was by Engle (2002), which can be attributed to the GARCH-X model category (Wang & Huang, 2012b). As the conditional variance in the EGARCH model is logarithmic, here we take the logarithm of the realized volatility and add it to the conditional variance equation with one lagging period. The corresponding model is called the EGARCH-X model, which is specifically expressed as follows:

$$\begin{cases} y_t = \mu + \delta y_{t-1} + \theta\varepsilon_{t-1} + \varepsilon_t, \\ \ln h_t = \omega + \beta ln h_{t-1} + \alpha\left|\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}}\right| + \gamma\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} + \varphi \ln RV_{t-1}, \\ \varepsilon_t|F_{t-1} \overset{iid}{\sim} \frac{2}{(\lambda+\lambda^{-1})\sqrt{h_t}}\left[f_v\left(\frac{\varepsilon_t}{\lambda\sqrt{h_t}}\right)\cdot I(\varepsilon_t \geq 0) + f_v\left(\frac{\varepsilon_t}{\lambda\sqrt{h_t}}\right)\cdot I(\varepsilon_t < 0)\right], \end{cases} \tag{9}$$

where $lnRV_{t-1}$ represents the logarithm of realized volatility at time t-1, and other symbol definitions are the same as those in the EGARCH model.

#### 3.1.3. The realGARCH model

Wang and Huang (2012a) pointed out that the GARCH-X model, in which the realized volatility is directly added to the variance equation, was not complete, because it cannot explain the changes in the realized volatility. To improve this point, researchers began to explore the 'complete' model, such as the relatively simple and straightforward Realized GARCH model (abbreviated as realGARCH) proposed by Hansen et al. (2012). In the model, a latent variable is used to realize the joint modelling of returns, volatility, and realized volatility, and the measurement equation includes the asymmetric response to shocks, that is, the leverage effect. Similar to the EGARCH and EGARCH-X models, the ARMA(1,1) processing is considered to model the conditional mean. The initial realGARCH model proposed by Hansen et al. (2012) has norm distribution residuals. Owing to the widespread 'fat tail' phenomenon in financial time series, Wang and Huang (2012a) extended the model to fat tail distribution. To examine the skewness of the residuals simultaneously, the residual distribution should to be sstd, still denoted as realGARCH, which is described as follows:

$$\begin{cases} y_t = \mu + \delta y_{t-1} + \theta\varepsilon_{t-1} + \varepsilon_t, \varepsilon_t = \sqrt{h_t}\cdot z_t, \\ \ln h_t = \omega + \beta ln h_{t-1} + \alpha \ln RV_{t-1}, \\ \ln RV_t = \xi + \phi \ln h_t + \tau(z_t) + u_t, u_t \overset{iid}{\sim} N\left(0,\ \sigma_u^2\right), \\ \varepsilon_t|F_{t-1} \overset{iid}{\sim} \frac{2}{(\lambda+\lambda^{-1})\sqrt{h_t}}\left[f_v\left(\frac{\varepsilon_t}{\lambda\sqrt{h_t}}\right)\cdot I(\varepsilon_t \geq 0) + f_v\left(\frac{\varepsilon_t}{\lambda\sqrt{h_t}}\right)\cdot I(\varepsilon_t < 0)\right], \end{cases} \tag{10}$$

where $h_t$ is called implied volatility. As $h_t$ and $RV_t$ are both measures of volatility, the third equation of (10), also called the measurement equation, is equivalent to taking $RV_t$ as a measured value of $h_t$. $\tau(z_t)$ represents leverage function and is set to a quadratic form as follows:

$$\tau(z_t) = \eta_1 z_t + \eta_2\left(z_t^2 - 1\right), \tag{11}$$

where $\eta_1$ is required to be negative, to capture asymmetric effects, and $E\tau(z_t) = 0$. The model of Eq. (10) is essentially the same as that of Wang and Huang (2012a).

### 3.2. Nonparametric models

#### 3.2.1. The NP model

The basic idea of GARCH modelling entails modelling the conditional mean and conditional variance of returns separately, and then assuming the distribution of residuals to obtain the distribution forecast of returns. However, is the hypothetical form of the model too strong? Will there be a bias in the model set? A natural idea is to directly model the distribution (or density) as a whole. Nonparametric researchers have proposed a series of prediction methods that directly target density and distribution based on the kernel method. Harvey and Oryshchenko (2012) systematically proposed a kernel-density estimation method for time series. For some time series, the cumulative distribution function forecast can be expressed as follows:

$$\hat{F}_{t+1|t}(y) = \omega\hat{F}_{t|t-1}(y) + (1-\omega)H\left(\frac{y-y_t}{h}\right), t = 1, 2, ..., T. \tag{12}$$

This is the recursive mechanism of the so-called NP model, which serves as a benchmark for nonparametric models. Where $\hat{F}(\cdot)$ represents the cumulative distribution function of the daily returns, $\mathrm{H}(\cdot)$ is a kernel function in the form of a cumulative distribution function, $\omega(0 \le \omega \le 1)$ is a decay factor, and h is a bandwidth. $\omega$, h are two parameters to be estimated, which can be realized by maximizing the average predicted log-likelihood function. The average predicted log-likelihood function is

$$\ell(\omega, \mathrm{h}) = \frac{1}{T-m}\sum_{t=m}^{T-1}\ln\hat{f}_{t+1|t}(y_{t+1}) = \frac{1}{T-m}\sum_{t=m}^{T-1}\ln\left[\frac{1}{h}\sum_{i=1}^{t}K\left(\frac{y_{t+1}-y_i}{h}\right)w_{t,i}(\omega)\right]. \tag{13}$$

Where $T$ is the total sample size and $m$ is a preset parameter. The practical processing is to use the initial $m$ observations to estimate $\hat{F}_{m|m-1}(y)$, and then use Eq. (13) to estimate the later $\hat{F}_{t|t-1}(y)$, thus the evaluation is only performed on the last $T-m$ distribution. $\hat{f}(\cdot)$ is the corresponding density function of the distribution function $\hat{F}(\cdot)$, and $\mathrm{K}(\cdot)$ is the kernel function of the density form corresponding to $\mathrm{H}(\cdot)$. In this paper, the Gaussian kernel function is used to describe $\mathrm{K}(\cdot)$, $m = 50$, and the initial value of the bandwidth $h$ is determined using a rule-of-thumb, namely,

$$\mathrm{h} = 1.05 \cdot \mathrm{S} \cdot T^{-\frac{1}{5}}, \tag{14}$$

Here S represents the standard deviation of the return time series. The exponential weighted moving average (EWMA) strategy is considered; as such, the weight function $w_{t,i}(\omega)$ is

$$w_{t,i}(\omega) = \frac{1-\omega}{1-\omega^t}\omega^{t-i}. \tag{15}$$

### 3.3.2. The realNP model

Based on Section 2.3, the distribution of $y_t$ derived from the intraday information is $y_t \sim N(N_t \cdot E(r_{t,i}),\ N_t \cdot Var(r_{t,i}))$, the cumulative distribution is denoted as $G_t(y)$, and the probability density function is denoted as $g_t(y)$. The EWMA strategy is still used, considering that the forecast distribution and density at time t + 1 are not only related to the daily return at time t but may also be related to the intraday high-frequency return at time t; therefore, the distribution forecast at time t + 1 is revised to

$$\hat{F}_{t+1|t}(y) = \omega \hat{F}_{t|t-1}(y) + (1-\omega)\left[\alpha H\left(\frac{y-y_t}{h}\right) + (1-\alpha)G_t(y)\right], t = 1, 2, ..., T. \tag{16}$$

The above formula is called the realNP model. Here $1-\alpha$ reflects the degree of interpretation of intraday information on the distribution of daily return, and $(1-\omega)\cdot(1-\alpha)$ reflects the influence of intraday information on the distribution forecast of the previous step. Therefore, the density forecast at time t + 1 is modified to

$$\hat{f}_{t+1|t}(y_{t+1}) = \sum_{i=1}^{t}\left[\alpha K_h\left(\frac{y_{t+1}-y_i}{h}\right) + (1-\alpha)g_i(y_{t+1})\right]w_{t,i}(\omega), \tag{17}$$

wherein $K_h\left(\frac{y_{t+1}-y_i}{h}\right) = \frac{1}{h}\mathrm{K}\left(\frac{y_{t+1}-y_i}{h}\right) = \frac{1}{\sqrt{2\pi}h}\exp\left[-\frac{(y_{t+1}-y_i)^2}{2h^2}\right]$. The average predicted log-likelihood function correspondingly becomes

$$\begin{aligned}\ell(\omega, \mathrm{h}) &= \frac{1}{T-m}\sum_{t=m}^{T-1}\ln\hat{f}_{t+1|t}(y_{t+1}) \\ &= \frac{1}{T-m}\sum_{t=m}^{T-1}\ln\left\{\sum_{i=1}^{t}\left[\alpha K_h\left(\frac{y_{t+1}-y_i}{h}\right) + (1-\alpha)g_i(y_{t+1})\right]w_{t,i}(\omega)\right\},\end{aligned} \tag{18}$$

wherein the calculation of $w_{t,i}(\omega)$ is consistent with the NP model.

## 4. Empirical results

This section presents the relevant empirical results from the full-sample fitting, the out-of-sample forecast of the rolling sample, the model combinations, and the economic evaluation.

### 4.1. Parameter estimation and full-sample fitting results

First, all empirical samples from 2010 to 2017 are used for the five models. The maximum likelihood estimation is used to solve each model. Table 2 presents the relevant parameter estimates. The EGARCH-X model directly adds the realized volatility into the variance equation of EGARCH. The $\varphi$ estimate of EGARCH-X is 0.1435 and is significant at the 1% level, and the log-likelihood function value of EGARCH-X has been raised from −2970.82 of EGARCH to −2953.04, indicating that the realized

**Table 2.** Parameter estimation and fitting of full sample in the empirical period.

| Model Parameter | EGARCH | EGARCH-X | realGARCH | NP | realNP |
|---|---|---|---|---|---|
| $\mu$ | 0.0189 | 0.0176 | 0.0043 | | |
| $\delta$ | −0.8648*** | −0.8563*** | 0.1618 | | |
| $\theta$ | 0.8782*** | 0.8735*** | −0.1788 | | |
| $\omega$ | 0.0022 | 0.0630*** | 0.1421*** | 0.9809 | 0.9751 |
| $\alpha$ | −0.0090 | −0.0254 | 0.2950*** | | 0.6592 |
| $\beta$ | 0.9946*** | 0.8541*** | 0.6882*** | | |
| $\gamma$ | 0.1199*** | 0.1564*** | | | |
| $\lambda$ | 0.9454*** | 0.9473*** | 0.9372*** | | |
| $\nu$ | 4.5900*** | 4.6200*** | 4.4978*** | | |
| $\varphi$ | | 0.1435*** | | | |
| $\xi$ | | | −0.4571*** | | |
| $\phi$ | | | 0.9914*** | | |
| $\sigma_u^2$ | | | 0.5231*** | | |
| $\eta_1$ | | | −0.1057* | | |
| $\eta_2$ | | | 0.9914*** | | |
| h | | | | 0.5002 | 0.2563 |
| LogL | −2970.82 | −2953.04 | −4447.24 | −2984.39 | −2939.38 |

**Note:** (1) *** represents significant at the 1% level. (2) The same parameter of different models may have different meanings, and the specific settings should be consistent with the previous model introduction. (3) LogL represents the log-likelihood function values.
Source: calculated by authors via R software.

volatility does affect the change of conditional variance. The scale-calibrated daily return distribution is derived from the intraday 5-minute return, and realNP directly adds the scale-calibrated distribution into NP. The $\alpha$ estimate of realNP is 0.6592, which means that the intraday 5-minute returns contribute $1-\alpha = 0.3408 = 34.08\%$ to the daily return distribution and only contribute $(1-\omega)\cdot(1-\alpha) = 0.85\%$ to the previous step distribution forecast. At the same time, for realNP, due to the integration of high-frequency information, the bandwidth h is significantly smaller, from 0.5002 of NP to 0.2563, and the log-likelihood function value increases from −2984.39 of NP to −2939.38. Comparisons of models EGARCH-X and EGARCH, realNP, and NP indicate that the intraday high-frequency return effectively improves the models' fitting effect. The intraday return is part of the description of the daily return distribution and the advance distribution forecast. The realGARCH model also includes intraday volatility; however, its log-likelihood function value is significantly lower than that of other models, which may be due to differences in its modelling mechanism, compared to other models.

Leverage effects are all considered in the parametric models. From Table 2, the $\gamma$ estimates of EGARCH and EGARCH-X are significantly positive and the $\eta_1$ estimate of realGARCH is significantly negative at the 1% level. This proves the existence of the leverage effect and indicates that the impact effect of negative returns on the next-day volatility is greater than that of positive returns.

Further comparison of NP and realNP indicates that the intraday return influences daily return distribution; however, the degree is less than expected. Its impact on the daily return distribution only accounts for $1-\alpha = 34.08\%$, while the 'single point' realization $y_t$ accounts for $\alpha = 65.92\%$. The impact of intraday information on the one-step-ahead distribution forecast is only $(1-\omega)\cdot(1-\alpha) = (1-0.9751)\cdot(1-0.6592) = 0.85\%$. This result may be attributed to two reasons: (1) the scale of intraday high-frequency returns is not properly calibrated, and (2) the intraday

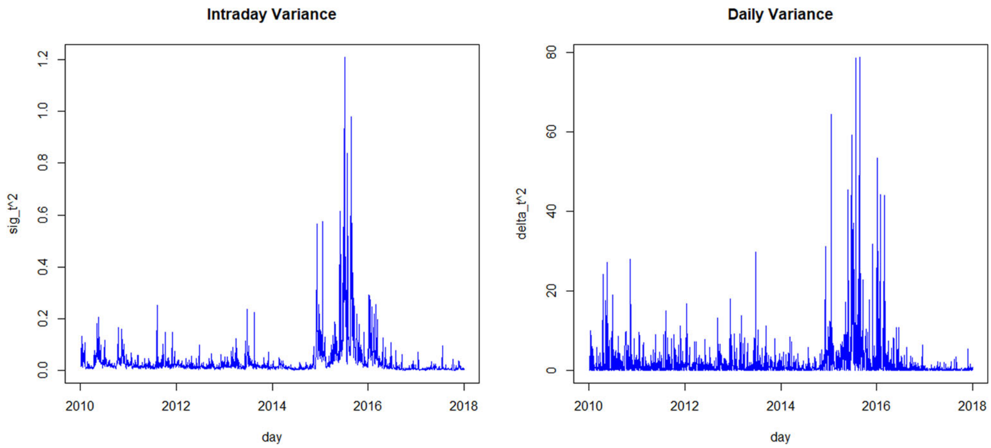


**Figure 5.** Intraday variance v.s. daily variance.
*Source:* compiled by authors with the help of R software.

returns are highly correlated with daily returns. Therefore, based on the $\hat{f}_{t+1|t}(y_{t+1})$ expression of Eq. (18), the 'single point' kernel density values of the daily return at time T is denoted as follows:

$$d_t = K_h\left(\frac{y_T - y_t}{h}\right), \ \ t = 1, 2, \ldots, T-1. \tag{19}$$

Then the intraday calibrated density values of the daily return at time T is denoted as follows:

$$din_t = g_t(y_T) = \frac{1}{\sqrt{2\pi Var(r_{t,i})}} \exp\left[-\frac{(y_T - \mu_{t,i})^2}{2Var(r_{t,i})}\right]. \tag{20}$$

The estimated result h = 0.2563 of realNP is substituted into the calculations of $d_t$ and $din_t$, then their Pearson correlation is calculated to obtain 0.6989, which means that the intraday return calibration effect is good, and the calibrated intraday return has a strong positive correlation with the daily return. In addition, the daily average of intraday returns is defined as $r_t = \frac{1}{N_t}\sum_{i=1}^{N_t} r_{t,i}$, and its variance is $\sigma_t^2 = \frac{1}{N_t - 1}\sum_{i=1}^{N_t}(r_{t,i} - r_t)^2$ (short form intraday variance). Thereafter, the variance of daily returns $\delta_t^2 = (y_t - \bar{y})^2$ (short form daily variance) is defined, where $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$. Figure 5 shows both the intraday and daily variances and that the two variance-series fluctuate similarly. The Pearson correlation of the two variances is 0.5367. Again, it shows that our intraday data calibration effect is good. Therefore, the intraday high-frequency return has a low impact on daily return distribution mainly because most of the intraday information ($din_t$) has been included in the 'single point' realization ($d_t$) of daily return $y_t$. Thus, for the distribution forecast of $y_t$, $din_t$ only complements $d_t$. This indicated the rationality of realNP modelling.

**Table 3.** Statistics of HL test in empirical period.

| Statistic Model | M(1, 1) | M(2, 2) | M(3, 3) | M(4, 4) | M(1, 2) | M(2, 1) | W |
|---|---|---|---|---|---|---|---|
| EGARCH | −0.89 | 1.06 | **2.12** | **2.66** | −1.46 | **2.98** | **12.36** |
| EGARCH-X | −1.32 | −0.44 | −0.10 | −0.04 | −1.57 | 0.88 | **11.81** |
| realGARCH | −2.00 | −1.44 | −0.89 | −0.62 | −2.16 | −0.30 | **5.80** |
| NP | −0.24 | **13.58** | **28.61** | **36.48** | **5.87** | **2.57** | **75.90** |
| realNP | 1.10 | **21.05** | **41.49** | **52.25** | **9.86** | **4.84** | **70.97** |

Source: calculated by authors via R software.

### 4.2. Statistical evaluation of distribution forecast

Section 4.1 shows the impact of high-frequency information on the fitting of the daily return distribution in a full sample. For both parametric and nonparametric methods, integrating high-frequency information helps to improve the goodness of fit; however, the impact of high-frequency information on the out-of-sample forecast may receive more attention. The following sections present three statistical evaluation results of the daily return distribution forecast. The three evaluations are probability integral transformation (PIT) evaluation, score evaluation, and evaluation of marginal calibration and sharpness.

#### 4.2.1. PIT evaluation

The years 2015, 2016, and 2017 are considered as the forecast interval, and the rolling window width is set as $T_0 = 1208$; thus, each model performs parameter estimation and one-step-ahead forecast $n = 732$ times. The return distribution forecast $\hat{F}^{(i)}_{t+1|t}(y)(i = 1, 2, \ldots, 5)$ can be obtained. As the realized value of the return at time $t + 1$ is $y_{t+1}$, the PIT sequence at time $t + 1$ is

$$PIT^{(i)}_{t+1} = \hat{F}^{(i)}_{t+1|t}(y_{t+1}). \tag{21}$$

If the model's forecast result can well reflect the data generation process, the PIT sequence should be independent and iid in the uniform distribution $U(0, 1)$, which is the probability calibration described in Gneiting et al. (2007).

There are two commonly used PIT tests: Berkowitz (2001)'s test (Berkowitz test for short) and Hong and Li (2005)'s test (HL test for short). If all models pass or fail the test, the Berkowitz test cannot be used to compare the advantages of the models. However, the HL test can be used to compare the quality of the models by the moment statistics M(i, j) and the W statistics; therefore, the HL test is adopted here. Taking the lag order as 4, Table 3 lists statistics for the HL test of the five models, and the significant values are emphasized in bold.

When the sample size is large enough, the statistics almost obey the standard normal distribution N(0, 1), and only the right-tail test is required. At the 5% significance level, the critical value of N(0, 1) is 1.65. Based on M(i, j) in Table 3, it can be concluded that the first four moments of EGARCH-X and realGARCH are correctly specified, but for the W statistic, the five models reject the null hypothesis, that is, the PIT sequence cannot be regarded as iid in $U(0, 1)$. Comparing the W statistic, it can be concluded that the order of the model from good to bad is realGARCH $>$

Table 4. Score and ranking of five competitive models in the empirical period.

| Model | Bayesian Winner | Winner ranking | Average Log-score | Log-score ranking | CRPS | CRPS ranking |
|---|---|---|---|---|---|---|
| EGARCH | 145 | 1 | −1.9809 | 4 | 0.8952 | 4 |
| EGARCH-X | 98 | 3 | −2.3626 | 5 | 1.0008 | 5 |
| realGARCH | 125 | 2 | −1.5063 | 1 | 0.7582 | 1 |
| NP | 81 | 5 | −1.8498 | 3 | 0.8697 | 3 |
| realNP | 91 | 4 | −1.8205 | 2 | 0.8609 | 2 |

Source: calculated by authors via R software.

$\mathrm{EGARCH}-\mathrm{X} > \mathrm{EGARCH} > \mathrm{realNP} > \mathrm{NP}$. Once again, this proves that, realGARCH and $\mathrm{EGARCH}-\mathrm{X}$ with high-frequency information are better than EGARCH, and realNP with high-frequency information, which in turn are better than NP. From a forecasting perspective, the realGARCH model is the best, and it realizes both realized volatility and 'implicit volatility' forecast modelling.

#### 4.2.2. Score evaluation

In many documents, the scoring rule of the logarithmic score is considered as correct and appropriate because of its many ideal attributes. The logarithmic score ranking of each model at time t is compared, and the model ranked first at time t is called the Bayesian winner. The more Bayesian winners are, the better the forecast performance of the model. In addition, the mean logarithmic score reflects the average similarity between the model forecasts and the observations. Furthermore, the Continuous Ranking Probability Score (CRPS) proposed by Gneiting et al. (2007), a scoring metric that considers sharpness, is also commonly used in the literature. Referring to Yao et al. (2020), this article adopts three scoring indicators: the number of Bayesian winners, average log score, and CRPS. Table 4 presents the relevant results. From the number and ranking of Bayesian winners, EGARCH is significantly better than other models, followed by realGARCH, EGARCH-X, realNP, and finally NP. From the average log score and CRPS, the order of the models from good to bad is realGARCH > realNP > NP > EGARCH > EGARCH-X. In general, realGARCH is the best model.

#### 4.2.3. Marginal calibration and sharpness

Marginal calibration and sharpness are distribution forecasting evaluation tools proposed by Gneiting et al. (2007) that do not rely on nested models. Gneiting et al. (2007) noted that the model should be calibrated for distribution prediction, especially for marginal calibration (Yao et al., 2021). Figure 6 shows the five models' marginal calibration diagrams. According to the calibration principle, the closer the calibration value is to 0, the better the calibration effect. As in Figure 6, for the two tails and the negative returns near zero, realGARCH, NP, and realNP have better calibration than EGARCH and EGARCH-X, and EGARCH is better than EGARCH-X. Within the $0\% - 1\%$ return range, NP and realNP show the worst calibration effects, and, overall, realGARCH has the best calibration effects. This may be due to the different modelling mechanisms and NP and realNP show great inconsistency with the parametric models. Parametric and nonparametric models exhibit 'non-synchronization' of marginal calibration. This inspired us to improve the model through a combination.

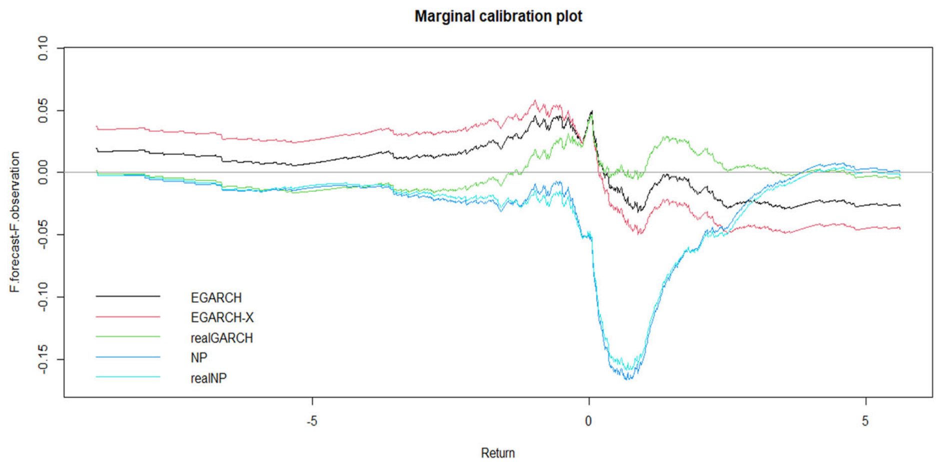


**Figure 6.** The marginal calibration plot during the empirical period.
*Source:* compiled by authors with the help of R software.

**Table 5.** Average width of forecasting intervals in the empirical period.

| Interval and ranking | EGARCH | EGARCH-X | realGARCH | NP | realNP |
|---|---|---|---|---|---|
| 50% interval | 2.1610 | 2.6525 | 1.5810 | 1.6778 | 1.6380 |
| 50% interval ranking | 4 | 5 | 1 | 3 | 2 |
| 90% interval | 4.9199 | 5.3761 | 4.3737 | 5.0320 | 4.8993 |
| 90% interval ranking | 3 | 5 | 1 | 4 | 2 |

Source: calculated by authors via R software.

Gneiting et al. (2007) noted that the distribution prediction be sharper after model calibration. The average widths of the 50% and 90% confidence intervals were used to examine the sharpness; the smaller the interval width, the better the sharpness. Table 5 presents the results, which indicate that for both 50% and 90% intervals, realGARCH is the best, and the performance of realNP and NP is relatively good.

### *4.3. Model combinations and economic evaluation*

As shown above, the marginal calibrations of different models are inconsistent, especially for the nonparametric models, which are different from the parametric models. Following Yao et al. (2020), this study considers three linear dynamic combination strategies: equal weight combination (EW), logarithmic score combination (SW), and CRPS combination (CW). Such combinations are all 'single-period' weighting, that is, the weight is calculated based on the forecast effect of one period only.

Next, an economic evaluation on the models is performed, as in Yao et al. (2020). First, based on the return distribution, this study derives two point forecasts of returns: mean and median, then compares the point forecasts with the realized values of return to calculate the direction accuracy (DA), furthermore the PT test of Pesaran and Timmermann (1992) is conducted to verify whether the returns are predictable. Second, a simple simulated trading strategy is examined based on the direction forecast under the short-selling situation, in which a virtual investor can sell short if the forecast for the next period of return is positive. In that case, the trading strategy

**Table 6.** Empirical direction accuracy and excess profitability of forecasts (unit: %).

| | Mean | | | Median | | |
|---|---|---|---|---|---|---|
| Model | DA | $Rate^{(1)}$ | $Rate^{(2)}$ | DA | $Rate^{(1)}$ | $Rate^{(2)}$ |
| EGARCH | 49.32 | 4.60 | 4.88 | 52.05 | 5.64** | 5.92 |
| EGARCH-X | 49.04 | −1.70 | −1.56 | 51.64 | 2.18 | 2.46 |
| realGARCH | 48.63 | −1.76 | −1.46 | 51.78 | 0.99 | 1.28 |
| NP | 56.56 | 0.29 | 0.58 | 56.56 | 0.29 | 0.58 |
| realNP | 56.56 | 0.29 | 0.58 | 56.56 | 0.29 | 0.58 |
| EW | 54.92 | −1.40 | −1.11 | 56.14 | 1.20 | 1.49 |
| SW | 59.42*** | 19.91*** | 20.14 | 56.97* | 6.64*** | 6.92 |
| CW | 58.20*** | 7.71*** | 7.98 | 56.42 | 2.03* | 2.32 |

**Notes:** (1) *, **, and *** represent significant results at the 10%, 5%, and 1% confidence levels respectively. (2) The ideal MTR in the short-selling case is 103.34%, and the ideal MTR in the non-short-selling case is 51.82%.
Source: calculated by authors via R software.

issues a buy signal; otherwise, it is a sell signal. Further, the ratio of the strategy mean transaction return (MTR) to the ideal MTR, that is $Rate^{(1)}$, is calculated as in Yao et al. (2020), and the EP test of Anatolyev and Gerko (2005) is conducted to test excess profitability. Finally, according to the actual situation of China's stock market, the ratio of the strategy MTR to the ideal MTR is examined under the non-short-selling situation, that is $Rate^{(2)}$, as in Yao et al. (2020). Table 6 shows the DA and economic evaluation results of individual models and combined models. From Table 6, all individual models and the EW model do not have significant predictability and excess profitability, except EGARCH. However, the combined models: SW and CW, behave differently. In the median case, the CW model shows excess profitability at the 10% significance level. The SW model performs best, whether for mean forecast or median forecast, and it shows direction predictability and significant excess profitability in the short-selling situation. For non-short-selling, no suitable test for excess profitability is employed; however, $Rate^{(2)} > Rate^{(1)}$ always holds. Therefore, we conclude that in the non-short-selling situation, the excess profitability of SW and CW should also be significant.

## 5. Robustness test

The robustness test was performed from 4 January 2010 to 30 June 2021. The out-of-sample prediction interval was from 2 January 2018 to 30 June 2021. The robustness test includes the models' parameter estimations—especially the parameter estimation of the realNP model—the statistical distribution forecasting evaluation of the five individual models, and the model combination and economic evaluation.

The parameters of the five individual models were estimated using data from 4 January 2010 to 30 June 2021, and the results are shown in Table 7. A comparison between Table 2 and Table 7 shows a small difference between the two. Except for $\delta$ and $\theta$ of realGARCH, other parameter estimates have the same signs. Regarding significance, $\delta$ and $\theta$ of realGARCH changed from insignificant in Table 2 to significant in Table 7. The significance of $\gamma$ of EGARCH-X changed from 1% significance in Table 2 to 5% significance in Table 7. Furthermore, we can draw several conclusions similar to the empirical period, such as EGARCH-X with intraday information fits better than EGARCH; realNP with intraday information fits better than NP; and

**Table 7.** Parameter estimation and fitting of full-sample in the robustness test period.

| Model Parameter | EGARCH | EGARCH-X | realGARCH | NP | realNP |
|---|---|---|---|---|---|
| $\mu$ | 0.0224 | 0.0181 | 0.0055 | | |
| $\delta$ | −0.9085*** | −0.9061*** | −0.9170*** | | |
| $\theta$ | 0.9221*** | 0.9226*** | 0.9235*** | | |
| $\omega$ | 0.0034 | 0.1399*** | 0.1551*** | 0.9827 | 0.9766 |
| $\alpha$ | −0.0111 | −0.0317 | 0.3103*** | | 0.6506 |
| $\beta$ | 0.9910*** | 0.6943*** | 0.6684*** | | |
| $\gamma$ | 0.1334*** | 0.1122** | | | |
| $\lambda$ | 0.9531*** | 0.9556*** | 0.9438*** | | |
| $\nu$ | 4.4634*** | 4.5712*** | 4.6276*** | | |
| $\varphi$ | | 0.2788*** | | | |
| $\xi$ | | | −0.4707*** | | |
| $\phi$ | | | 0.9793*** | | |
| $\sigma_u^2$ | | | 0.5295*** | | |
| $\eta_1$ | | | −0.1005*** | | |
| $\eta_2$ | | | 0.1237 *** | | |
| h | | | | 0.5214 | 0.2701 |
| LogL | −4248.20 | −4212.05 | −6407.13 | −4308.36 | −4250.45 |

**Note:** ** and *** represent significant at the 5% and 1% levels, respectively.
Source: calculated by authors via R software.

**Table 8.** Score and ranking of five competitive models in the robustness test period.

| Model | Bayesian Winner | Winner ranking | Average Log-score | Log-score ranking | CRPS | CRPS ranking |
|---|---|---|---|---|---|---|
| EGARCH | 130 | 4 | 1.5071 | 2 | 0.6256 | 2 |
| EGARCH-X | 98 | 5 | −1.5200 | 3 | 0.6415 | 4 |
| realGARCH | 145 | 3 | −1.4915 | 1 | 0.6214 | 1 |
| NP | 163 | 2 | −1.7317 | 5 | 0.6450 | 5 |
| realNP | 312 | 1 | −1.7016 | 4 | 0.6395 | 3 |

Source: calculated by authors via R software.

intraday information improves the goodness-of-fit. In realNP, intraday information explains $1-\alpha = 34.94\%$ of the current day's return and contributes $(1-\omega)\cdot(1-\alpha) = (1-0.9766)\cdot(1-0.6506) = 0.82\%$ to the one-step-ahead return distribution, which is also consistent with the conclusion of the empirical period.

Considering the generality of the score evaluation and its relation with the model combination, the robustness test of the statistical forecast evaluation is only conducted from the score evaluation. Table 8 shows the score evaluation results of the daily return distribution forecast from 2018 to 2021. Comparing Table 4 and Table 8, for the Bayesian winner, the nonparametric models perform worse than the parametric models during the empirical period; however, in the robustness test period, the nonparametric models are significantly better than the parametric models. Regarding average log score and CRPS, the nonparametric models are, overall, inferior to the parametric models during the robustness test period, which differs from the results in the empirical period. The above-mentioned 'non robustness' confirms the necessity of joint modelling of parametric and nonparametric models. However, some commonalities also exist. Overall, it realGARCH is considered as relatively optimal. Regardless of the kind of evaluation, realNP is always better than NP.

Similar to Section 4.3, the model combination and economic evaluation are investigated in the robustness test period, and the relevant results are shown in Table 9.

**Table 9.** Direction accuracy & excess profitability of forecasts for robustness test (unit: %).

| | Mean | | | Median | | |
|---|---|---|---|---|---|---|
| Model | DA | $Rate^{(1)}$ | $Rate^{(2)}$ | DA | $Rate^{(1)}$ | $Rate^{(2)}$ |
| EGARCH | 50.83 | 5.05 | 6.12 | 51.30 | 1.89 | 3.00 |
| EGARCH-X | 52.00 | 7.49* | 8.54 | 51.65 | 3.94** | 5.02 |
| realGARCH | 51.06 | 5.03 | 6.11 | 50.83 | −3.03 | −1.87 |
| NP | 48.35 | −1.14 | 0.21 | 51.56 | 1.20 | 2.26 |
| realNP | 48.55 | −1.04 | 0.25 | 51.65 | 1.34 | 2.42 |
| EW | 50.71 | 9.06** | 10.09 | 51.77 | 2.23 | 3.34 |
| SW | 51.06 | 7.47* | 8.52 | 51.42 | 1.21 | 2.32 |
| CW | 52.00 | 12.21*** | 13.20 | 51.77 | 2.57** | 3.67 |

**Notes**: (1) *, **, and *** represent significant results at the 10%, 5%, and 1% confidence levels, respectively. (2) The ideal MTR in the short-selling case is 85.06%, and the ideal MTR in the non-short-selling case is 43.02%.
Source: calculated by authors via R software.

From Table 9, all individual models and combined models have no significant direction predictability, but the DA of combined models is generally better than that of individual models, which is consistent with the results during the empirical period reflected in Table 6. During the robustness test period, SW, the best-performing combination during the empirical period, exhibits excess profitability only in the mean case, at the 10% level, but it is not as good as EW. Further, CW exhibits significant excess profitability, both in the mean and median cases. In addition, the excess profit in the non-short-selling situation all exceeds the short-selling situation, which is consistent with the conclusion in the empirical period, meaning the non-short-selling rule may have a protective effect on investors.

## 6. Conclusions

The daily return distribution forecast with intraday trading information is examined, using the Shanghai Composite Index in China's stock market as the research object. The intraday information is reflected by the 5-minute closing price and is integrated into the daily return model using two methods: realized volatility and intraday scale calibrated return. From both parametric and nonparametric perspectives, the quantitative impact of intraday high-frequency information on daily returns is studied through model comparison.

The full sample fitting results show that, (1) the intraday high-frequency information improves the goodness-of-fit. (2) A comparison of the NP and realNP models shows that the intraday return can only contribute about a 30% description of the daily distribution, and it can only provide less than 1% of the information for the one-step-ahead distribution forecast. This is attributed to the scale calibrated intraday return that is closely related to the daily return and it reflects that our scale calibration of intraday returns is good.

The statistical evaluation of out-of-sample forecasting shows that, (1) high-frequency information helps to improve the forecasting effect. (2) Under different evaluation criteria, the model ranking is different. Overall, the realGARCH model is relatively the best. The inconsistency in the performance of parametric and nonparametric models implies the necessity for combinatorial modelling.

The implementation of three model combination strategies: EW, SW, and CW, as well as economic evaluation reveal that individual models often do not have significant predictability and excess profitability; however, the combined models behave differently, that is, the three combinations sometimes demonstrate directional predictability and excess profitability. SW and CW perform best in the empirical and robustness test periods, respectively. The excess profit in the non-short-selling situation is always more than that in a short-selling situation, which implies that the non-short-selling rule has a protective effect on investors.

This study only researches China's stock market; however, it can provide references to the forecast modelling of the stock market in other countries, mainly in the scale calibration, model comparison, and model combination methods. This study still has some shortcomings, for instance, (1) there are many ways to measure realized volatility, but this study only uses the most traditional one, and does not make other attempts and comparisons, such as the generalized realized measures proposed by Jiang et al. (2018) and Jiang and Gu (2019). (2) The difference in the processing mechanism the deduced intraday scale calibration effect cannot be compared with that of Hallam and Olmo (2014a, 2014b). (3) The modelling of realNP lacks rigorous mathematical proof, and (4) no suitable method is selected to conduct a statistical test on the excess profitability of the non-short-selling situation.

## Disclosure statement

No potential conflict of interest was reported by the author.

## Funding

This work was supported by the Zhejiang Province Philosophy and Social Science Planning Project under Grant No. 21NDJC172YB; and Major Humanities and Social Sciences Research Youth Key Projects in Zhejiang Colleges and Universities under Grant No. 2021QN014.

## Notes on contributors

*Yanyun Yao*, is an associate professor at Ningbo University of Finance & Economics and a part-time master tutor at Shaoxing University. Her main research field is econometrics.

*Qiang Huang*, is a master's degree student at Shaoxing University. His main research field is financial mathematics.

*Shangzhen Cai*, is a lecturer at Ningbo University of Finance & Economics. His main research field is computer application.

## References

Anatolyev, S., & Gerko, A. (2005). A trading approach to testing for predictability. *Journal of Business & Economic Statistics*, *23*(4), 455–461. https://doi.org/10.1198/073500104000000640

Andersen, T. G., & Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, *4*(2-3), 115–158. https://doi.org/10.1016/S0927-5398(97)00004-2

Andersen, T. G., & Bollerslev, T. (1998). Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. *The Journal of Finance*, *53*(1), 219–265. https://doi.org/10.1111/0022-1082.85732

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, *96*(453), 42–55. https://doi.org/10.1198/016214501750332965

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2003). Modeling and forecasting realized volatility. *Econometrica*, *7*(2), 579–625.

Bandi, F. M., & Russell, J. R. (2008). Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies*, *75*(2), 339–369. https://doi.org/10.1111/j.1467-937X.2008.00474.x

Barndorff-Nielsen, O. E., & Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Economics*, *2*(1), 1–37.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2008). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Economometrica*, *76*(6), 1481–1536.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, *19*(4), 465–474. https://doi.org/10.1198/07350010152596718

Brooks, C., Burke, S., & Heravi, S. (2005). Autoregressive conditional kurtosis. *Journal of Financial Econometrics*, *3*(3), 399–421. https://doi.org/10.1093/jjfinec/nbi018

Clements, M. P., Galvao, A., & Kim, J. (2008). Quantile forecasts of daily exchange rate returns from forecasts of realized volatility. *Journal of Empirical Finance*, *15*(4), 729–750. https://doi.org/10.1016/j.jempfin.2007.12.001

Engle, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics*, *17*(5), 425–446. https://doi.org/10.1002/jae.683

Engle, R. F., & Sokalska, M. E. (2012). Forecasting intraday volatility in the US equity market: Multiplicative component garch. *Journal of Financial Econometrics*, *10*(1), 54–83. https://doi.org/10.1093/jjfinec/nbr005

Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration, and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *69*(2), 243–268. https://doi.org/10.1111/j.1467-9868.2007.00587.x

Gorgi, P., Koopman, S. J., & Li, M. (2019). Forecasting economic time series using score-driven dynamic models with mixed-data sampling. *International Journal of Forecasting*, *35*(4), 1735–1747. https://doi.org/10.1016/j.ijforecast.2018.11.005

Hallam, M., & Olmo, J. (2014a). Semiparametric density forecasts of daily financial returns from intraday data. *Journal of Financial Econometrics*, *12*(2), 408–432. https://doi.org/10.1093/jjfinec/nbt016

Hallam, M., & Olmo, J. (2014b). Forecasting daily return densities from intraday data: A multifractal approach. *International Journal of Forecasting*, *30*(4), 863–881. https://doi.org/10.1016/j.ijforecast.2014.01.007

Hansen, P. R., & Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, *24*(2), 127–161. https://doi.org/10.1198/073500106000000071

Hansen, P. R., Huang, Z., & Shek, H. H. (2012). Realized garch: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, *27*(6), 877–906. https://doi.org/10.1002/jae.1234

Harvey, A., & Oryshchenko, V. (2012). Kernel density estimation for time series data. *International Journal of Forecasting*, *28*(1), 3–14. https://doi.org/10.1016/j.ijforecast.2011.02.016

Hong, Y., & Li, H. (2005). Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *Review of Financial Studies*, *18*(1), 37–84. https://doi.org/10.1093/rfs/hhh006

Huang, Y., Tang, Z., & Zhou, X. (2015). Estimation of tail risk based on realized GARCH model with skew-t distribution. *Systems Engineering - Theory & Practice (in Chinese)*, *35*(09), 2200–2208.

Huffman, S. P., & Moll, C. R. (2012). The impact of asymmetry on expected stock returns: an investment of alternative risk measures. *Algorithmic Finance*, *1*(2), 79–93. https://doi.org/10.3233/AF-2011-008

Hwan, K. D., Eun, M. S., & Sik, B. Y. (2011). Effects of intraday patterns on a time series analysis of the Korean stock market index. *Journal- Korean Physical Society*, *58*(2), 396–399.

Jiang, W., & Gu, Y. (2019). Forecasting VaR: Realized GARCH Incorporating Generalized Realized Measures. *The Journal of Quantitative & Technical Economics (in Chinese)*, (7), 156–173.

Jiang, W., Ruan, Q., Li, J., & Li, Y. (2018). Modeling returns volatility: Realized GARCH incorporating realized risk measure. *Physica A: Statistical Mechanics and Its Applications*, *500*, 249–258. https://doi.org/10.1016/j.physa.2018.02.018

Lu, B., Ma, F., Wang, J., Ding, H., & Wahab, M. I. M. (2021). Harnessing the decomposed realized measures for volatility forecasting: Evidence from the US stock market. *International Review of Economics & Finance*, *72*, 672–689. https://doi.org/10.1016/j.iref.2020.12.023

Lyócsa, Š., Molnár, P., & Výrost, T. (2021). Stock market volatility forecasting: Do we need high-frequency data? *International Journal of Forecasting*, *37*(3), 1092–1110. https://doi.org/10.1016/j.ijforecast.2020.12.001

Maheu, J. M., & McCurdy, T. H. (2011). Do high-frequency measures of volatility improve forecasts of return distributions? *Journal of Econometrics*, *160*(1), 69–76. https://doi.org/10.1016/j.jeconom.2010.03.016

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, *59*(2), 347–370. https://doi.org/10.2307/2938260

Pesaran, M. H., & Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics*, *10*(4), 461–465.

Tsay, R. S. (2010). *Analysis of Financial Time Series (Third.)*. John Wiley.

Wang, C., & Nishiyama, Y. (2015). Volatility forecast of stock indices by model averaging using high-frequency data. *International Review of Economics & Finance*, *40*, 324–337. https://doi.org/10.1016/j.iref.2015.02.014

Wang, T., & Huang, Z. (2012a). High-frequency volatility modeling. *Quantitative Economics and Technical Economic Research (in Chinese)*, *29*(5), 149–161.

Wang, T., & Huang, Z. (2012b). Review of research on volatility modeling and application based on high-frequency data. *Economic Dynamics (in Chinese)*, (03), 141–146.

Wang, T., & Huang, Z. (2015). Realized GAS-GARCH and its application in VaR prediction. *Journal of Management Sciences in China (in Chinese)*, *18*(05), 79–86.

Wuertz, D., Chalabi, Y., & Chen, W. (2009). *Portfolio optimisation with r/rmetrics, rmetrics eBook*. Rmetrics Association and Finance Online.

Yao, Y., Cai, S., & Wang, H. (2021). Are technical indicators helpful to investors in China's stock market? A study based on some distribution forecast models and their combinations. *Economic Research-Ekonomska Istraživanja*, 1–25. https://doi.org/10.1080/1331677X.2021.1974921

Yao, Y., Zheng, X., & Wang, H. (2020). Predictability of China's stock market returns based on combination of distribution forecasting models. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, *24*(4), 477–486. https://doi.org/10.20965/jaciii.2020.p0477

Yu, X., & Wang, X. (2018). Mix frequency realized GARCH Models: The forecast of volatility and measure of VaR. *Statistical Research (in Chinese)*, *35*(01), 104–116.

Zhu, W., & Tang, D. (2017). Analysis of the influence of volatility high-frequency index on the forecasting ability of return distribution. *Chinese Market (in Chinese)*, (18), 63–64.