Mathematical and Computational Applications

Article

Volatility Forecasting Based on Cyclical Two-Component Model: Evidence from Chinese Futures Markets and Sector Stocks

Conghua Wen * and **Junwei Wei**

Department of Financial Mathematics, School of Science, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China; Junwei.Wei19@student.xjtlu.edu.cn

* Correspondence: conghua.wen@xjtlu.edu.cn; Tel.: +86-512-81884828

Received: 21 July 2020; Accepted: 9 September 2020; Published: 10 September 2020

Abstract: This article aims to study the schemes of forecasting the volatilities of Chinese futures markets and sector stocks. An improved method based on the cyclical two-component model (CTCM) introduced by Harris et al. in 2011 is provided. The performance of CTCM is compared with the benchmark model: Heterogeneous Autoregressive model of Realized Volatility type (HAR-RV type). The impact of open interest for futures market is included in HAR-RV type model. We employ 3 different evaluation rules to determine the most efficient models when the results of different evaluation rules are inconsistent. The empirical results show that CTCM is more accurate than HAR-RV type in both estimation and forecasting. The results also show that the realized range-based tripower volatility (RTV) is the most efficient estimator for both Chinese futures markets and sector stocks.

Keywords: high frequency; mean absolute percentage error; non-parametric filter; open interest; volatility forecasting

1. Introduction

The performance of financial markets is a keen signal for economic development worldwide [1]. From this respect, analyzing and forecasting the trend of financial market is crucial. In recent decades, volatility as the second moment structure of price is the most concerned by mathematicians and economists, due to its influence on assets pricing, risk management, portfolio construction and financial derivatives' pricing models such as Black-Scholes-Merton Model. Therefore, better measurement and model of volatility will contribute to more accurate pricing for academics and market investors, which is helpful for the different interested parties, for instance, managers, investors, and policy maker to adopt miscellaneous financial measures such as raising or reducing capital, investment allocation, etc.

Various factors impact on the effectiveness of volatility model [2]. One is the fitness of the proxy for unobserved volatility which is adopted in the model. Traditional volatility proxies based on the squared demeaned return are unbiased estimators of the latent integrated variance because the integrated volatility is, by construction, the expectation of the squared demeaned return. Nevertheless, proxies constructed by the squared return employ only a single measurement of the price each period and contain no information about the intra-period trajectory of the price, which causes inefficiency. An improvement in efficiency is using intraday data, such as 5-min high-frequency data. In this view, academics come up with several volatility estimators. Another important factor that influences the effectiveness of a volatility model is the specification of the process that governs volatility dynamics. Academics build various kinds of models concentrated on different characteristics, such as long memory [3], cycling jumps [4], autoregression [5], etc.

However, several volatility models failed to forecast immature markets, due to its sharp jumps and falls. Chinese market is still one of the emerging financial environments but plays a key role in global economics development. China's Gross Domestic Product has been the second largest since 2010. By the end of 2017, the total market value of the Shanghai and Shenzhen stock markets had reached more than 8 trillion U.S. dollars, ranking as the second largest in the world [6]. In March, 2018, Chinese crude oil futures came into the market, which increased the influence of Chinese futures markets. At the same time, an extreme market fluctuation appeared in 2015. The Shanghai Composite Index fell rapidly by 32% in the 17 trading days after June 12 with market capitalization down by a third [6]. Therefore, to obtain a model that works effectively in immature markets, it is of good values to investigate China's financial markets.

This article aims to forecast the volatilities of Chinese futures markets and sector stocks based on cycling model. As a comparison, Heterogeneous Autoregressive model of Realized Volatility (HAR-RV) type [5] originated by Corsi in 2009 and its improved models [7] introduced by Christensen K. et al. in 2012 are chosen as benchmark. To find a better volatility proxy, 3 realized ranged-based estimators are introduced. The empirical data sample is a set of 3 years 5-min high-frequency data downloaded from WIND database. The futures' sample are: silver, aluminum, copper traded in Shanghai Futures Exchange (SHFE); ironstone, coke, coking coal, soybean meal traded in Dalian Commodity Exchange (DCE); and rapeseed meal traded in Zhengzhou Commodity Exchange (CZCE). The sector stocks' data sample are energy index, raw materials index, medical hygiene index, and financial real estate index listed in Shanghai Stock Exchange.

Volatility problem initiated from random walk. In 1980s, it was generally accepted that the price of a common stock followed a random walk. The volatility thus became a key variable to calculate. Parkinson [8] provided a method to estimate the volatility called realized range (volatility) (RRV), which was shown that 2–5 times better than squared return, the traditional volatility, in 1980. This method used the scaled difference of intraday highest and lowest price of stock to measure the volatility. Recently, RRV attracted attention again. Some literature [9] showed that it was not only significantly more efficient than the squared return, but also more robust than realized volatility, another volatility estimator provided by Andersen and Bollerslev in 1998 [10] to market microstructure noise.

In 2012, Christensen and Podolskij [7] based on the multipower variance measures introduced a complete realized range-based multipower variance (RMV) approach. Rather than employing the variance, RMV calculates volatility estimator with absolute returns in ranges. RMV provides considerable efficient results even though the high-frequency data provide sparse information. RMV provides an asymptotic efficient estimation, which is also robust to jumps. Another advantage of RMV is outstanding in the presence of market microstructure noise. Evidence [7] shows that higher-order RMV will usually get better estimation behavior; however, the third order is efficient enough in practice. Therefore, this paper employs realized range-based bipower volatility (RBV) and realized range-based tripower volatility (RTV) to model Chinese data.

Besides the development of various volatility estimators, academics also paid attention to the volatility models themselves. Muller et al. [11] presented the presence of heterogeneity across investors in 1993. In 1997, Andersen and Bollerslev [12] claimed the volatility process also contained multiple components. Corsi [5] concentrated on the heterogeneity that originates from the difference in the trading frequency and introduced HAR-RV model in 2009. Corsi separated the traders into 3 main parts: daily traders, weekly traders, and monthly traders. HAR-RV model led to a simple AR-type model was considered to have different volatility components based on time horizons, which had a good estimation and forecasting performance. In 2012, Christensen and Podolskij [7] provided a realized range-based multipower variation approach, which eliminated the bias by constructing a hybrid range-based estimator. Replacing the weekly and monthly RV by RBV or RTV, the model called HAR-RRV-RMV model revealed the better efficiency than traditional HAR-RV model.

Some academics suggested that volatility could be divided into two components, processes governing long-term or short-term dynamics [4]. In 1998, Engle and Lee [13] originated a component

GARCH model, which separated volatility into a constant long-run trend component and a temporary short-run component, i.e., mean reverting towards the long-run trend. Empirical evidence [2,9] revealed that the two-factor model had better performance than a one-factor model. The forecasting horizon of two-factor model is up to 1 year [2]. Harris et al. [4] provided a cyclical two-component model to estimate and forecast volatility over both short and long horizons. The long-term trend volatility was estimated by a non-parametric filter, while the short-term component was modelled by a stationary AR process based on the long-term component. The results gave reliable estimation and forecasting, compared with one-factor and two-factor range-based EGARCH model and the range-based FIEGARCH model [2] introduced by Brandt and Jones in 2006.

Motivated by Tianlun [14] who tried to find a reasonable volatility forecasting method for Chinese individual stocks by comparing 6 volatility estimators applied in 3 models, this paper employs 3 estimators in 2 models to estimate and forecast volatility of Chinese futures market and sector stocks. The estimators contain HRV, HBV, and HTV. While the models cover HRV-RV type and 3 kinds of cyclical two-component models (CTCM).

The contributions of this paper are concluded in 3 main points.

First, this paper is the first paper to comprehensively study the volatility models of Chinese futures markets and sector stocks. The volatility estimators and volatility models are popular and show good predictability for recent years. This paper also considers the impact of Open Interest, which is important but less studied.

Secondly, this paper finds a common model to estimate Chinese futures and sector stocks, no matter the distributions and fluctuations of the products. Improved CTCM introduced in this paper shows robust estimation results.

Finally, this article employs different evaluation rules and provides an idea to determine the most efficient model when the evaluation rules are not consistent.

The remainder of this paper is separated in 4 parts. Section 2 is the methodology; Section 3 is the empirical results; Section 4 is the discussion; and Section 5 concludes the study.

2. Data and Methods

2.1. Data

This paper employs the data downloaded from WIND database including around 1000 day 5-min interval intraday data for the period from 14 December 2015 to 11 August 2020, totally 4.75-year data. The data contains date, time, opening price, closing price, highest price, lowest price and open interest (for futures only), which is used to calculate the volatility estimators and construct volatility models. The futures' trading time is from 9:00 a.m. to 11:30 a.m., 1:30 p.m. to 3:00 p.m. and 9:00 p.m. to the next day 2:30 a.m., while the sector stock's trading time is from 9:30 a.m. to 11:30 a.m. and 1:00 p.m. to 3:00 p.m. In particular, if the data in some time are missing, it will be replaced by the same data as the last time.

To avoid the interplay of upstream-downstream industries, data from different areas are selected as experimental sample. The metals area contains silver (code: AG.SHF, hereafter, Ag(SHFE)), aluminum (code: AL.SHF, hereafter Al(SHFE)), copper (code: CU.SHF, hereafter Cu(SHFE)) and ironstone (code: I.DEC, hereafter I(DEC)). The non-renewable energy resource area contains coke (code: J.DCE, hereafter J(DCE)) and coking coal (code: JM.DCE, hereafter JM(DCE)). The agricultural and sideline products area contains soybean meal (code: M.DCE, hereafter M(DCE)) and rapeseed meal (code: RM.CZC, hereafter RM(CZCE)). The sector stocks' sample contains energy index (code: 000032.SH, hereafter ENG), raw materials index (code: 000033.SH, hereafter MTR), medical hygiene index (code: 000037.SH, hereafter MDC), and financial real estate index (code: 000038.SH, hereafter FINRE).

The efficiency of model depends on sample size, thus most of articles used 10-year high-frequency data to investigate the effectiveness of models, and set the forecast period as 1 year. Due to the limit of practice, the database provides 5-min data covering only latest 3 years in China. Our data start from 14 December 2015 and end on 11 August 2020, around 1000 days in total. During forecasting, this paper employs 124 days data as out of sample forecasting period and other days as estimation period to construct models. There are two reasons for such a choice. First, the futures contracts usually have half-a-year holding period. Secondly, most papers employed 9-year data to construct model and 1-year data to forecast due to the evidence that the volatility estimators have 1-year predictability, therefore the predictive proportion is 1 over 10. Hence, we choose predictive and training proportions as around 1 over 9.

Before constructing models, the summary statistics of volatility estimators for each product (Tables 1–3) are shown below, containing minimum, maximum, median, mean, variance, kurtosis, and skewness. These numbers describe the distribution and the fluctuation of estimators, which may be a cause of the estimation's or the model's poor performance.

Table 1. Summary statistics of RRV for each product.

	Ag	**Al**	**Cu**	**I**	**J**	**JM**	**M**	**RM**	**ENG**	**MTR**	**MDC**	**FINRE**
Minimum	0.0016	0.0023	0.0028	0.0168	0.0066	0.0123	0.0039	0.0044	0.0049	0.0042	0.0019	0.0033
Maximum	0.4644	0.5194	0.6362	1.0039	0.8519	0.9947	0.3985	0.6752	0.1428	0.1352	0.1556	0.1510
Median	0.0127	0.0179	0.0154	0.1292	0.0953	0.0944	0.0230	0.0355	0.0140	0.0131	0.0131	0.0117
Mean	0.0240	0.0250	0.0223	0.1464	0.1210	0.1291	0.0306	0.0461	0.0177	0.0180	0.0179	0.0157
Variance	0.0013	0.0007	0.0010	0.0086	0.0082	0.0129	0.0008	0.0016	0.0002	0.0002	0.0003	0.0002
Kurtosis	37.190	105.47	158.89	14.420	13.082	13.204	38.401	65.167	20.271	13.469	10.513	21.577
Skewness	5.1229	7.3240	10.348	2.7295	2.7589	2.8546	3.6390	5.4334	3.6420	2.9280	2.6337	3.6954

Table 2. Summary statistics of RBV for each product.

	Ag	**Al**	**Cu**	**I**	**J**	**JM**	**M**	**RM**	**ENG**	**MTR**	**MDC**	**FINRE**
Minimum	0.0013	0.0021	0.0021	0.0155	0.0105	0.0118	0.0035	0.0040	0.0050	0.0042	0.0019	0.0033
Maximum	0.3612	0.4923	0.5504	0.8902	0.7433	0.9807	0.3179	0.5522	0.1226	0.0941	0.1308	0.1263
Median	0.0114	0.0158	0.0136	0.1183	0.0825	0.0829	0.0205	0.0305	0.0133	0.0121	0.0109	0.0106
Mean	0.0205	0.0220	0.0192	0.1322	0.1064	0.1155	0.0266	0.0396	0.0164	0.0160	0.0151	0.0139
Variance	0.0009	0.0006	0.0006	0.0065	0.0063	0.0105	0.0006	0.0011	0.0001	0.0001	0.0002	0.0001
Kurtosis	35.256	134.73	214.82	13.532	13.568	13.429	31.000	60.914	20.154	8.0308	10.137	20.694
Skewness	5.0448	8.2504	11.358	2.5689	2.7784	2.8763	4.0837	5.1672	3.5481	2.4790	2.6025	3.5600

Table 3. Summary statistics of RTV for each product.

	Ag	**Al**	**Cu**	**I**	**J**	**JM**	**M**	**RM**	**ENG**	**MTR**	**MDC**	**FINRE**
Minimum	0.0012	0.0020	0.0016	0.0150	0.0099	0.0116	0.0031	0.0037	0.0050	0.0042	0.0018	0.0032
Maximum	0.2491	0.4644	0.4998	0.8474	0.7100	0.9538	0.2756	0.4952	0.1104	0.0855	0.1199	0.1128
Median	0.0107	0.0148	0.0128	0.1129	0.0775	0.0790	0.0191	0.0286	0.0130	0.0117	0.0100	0.0100
Mean	0.0190	0.0207	0.0181	0.1260	0.1003	0.1096	0.0250	0.0369	0.0158	0.0152	0.0138	0.0132
Variance	0.0007	0.0005	0.0005	0.0059	0.0056	0.0095	0.0005	0.0009	0.0001	0.0001	0.0002	0.0001
Kurtosis	26.834	138.98	199.71	13.639	13.078	13.355	25.989	54.450	18.898	7.3582	10.270	18.801
Skewness	4.5599	8.3728	10.922	2.5773	2.7257	2.8670	3.7829	4.8830	3.4398	2.3910	2.6114	3.4223

2.2. Volatility Estimators

This part introduces equations of volatility estimators, containing RRV, RBV and RTV.

Let H_t denote the highest price at an intraday time t, L_t denote the lowest price at the same time t, and T denote the amount of an intraday data. The Lambda function is defined as

$$\lambda(r) = \frac{4}{\sqrt{\pi}}\left(1 - \frac{4}{2^r}\right) 2^{r/2}\Gamma((r+1)/2)\zeta(r-1), \tag{1}$$

where $\Gamma(\cdot)$ is Gamma function, and $\zeta(\cdot)$ is Zeta function. In particular, $\lambda(2) = 4\log 2$.

Then the RRV_i representing the ith day volatility estimator is defined as

$$RRV_i = \frac{1}{\lambda(2)} \sum_{t=1}^{T} (\log H_{t,i} - \log L_{t,i})^2, \tag{2}$$

where $t = 1, 2, 3, \cdots, T$ is the order of the intraday price.

The RBV_i representing the ith day volatility estimator is defined as

$$RBV_i = \frac{T}{T-1} \left(\frac{1}{\lambda(1)}\right)^2 \sum_{t=2}^{T} (\log H_{t,i} - \log L_{t,i})(\log H_{t-1,i} - \log L_{t-1,i}), \tag{3}$$

where $t = 1, 2, 3, \cdots, T$ is the order of the intraday price.

The RTV_i representing the ith day volatility estimator is defined as

$$RTV_i = \tfrac{T}{T-2} \left(\tfrac{1}{\lambda(2/3)}\right)^3 \sum_{t=3}^{T} (\log H_{t,i} - \log L_{t,i})^{\frac{2}{3}} (\log H_{t-1,i} - \log L_{t-1,i})^{\frac{2}{3}} (\log H_{t-2,i} - \log L_{t-2,i})^{\frac{2}{3}}, \tag{4}$$

where $t = 1, 2, 3, \cdots, T$ is the order of the intraday price.

2.3. Volatility Models

This part introduces the volatility models including HAR-RV type and CTCM type.

2.3.1. HAR-RV Type Models

The basic HAR-RV type model is HAR-RV. This model only considers the relation between estimator and its past moving average. Let $\hat{\sigma}_t$ denote the volatility estimator, $\bar{\sigma}_{n,t}$ denote the nth moving average at time t which is calculated by

$$\bar{\sigma}_{n,t} = (\hat{\sigma}_t + \hat{\sigma}_{t-1} + \cdots + \hat{\sigma}_{t-n+1})/n. \tag{5}$$

Noting that $n = 5$ represents weekly data and $n = 22$ represents monthly data. Then HAR-RV model is defined by

$$\hat{\sigma}_t = \beta + \beta_d \hat{\sigma}_{t-1} + \beta_w \bar{\sigma}_{5,t-1} + \beta_m \bar{\sigma}_{22,t-1} + \epsilon_t, \tag{6}$$

where β is constant term, β_d, β_w, β_m are the coefficients of daily term, weekly term and monthly term, respectively, and ϵ_t is the residual at time t.

Besides the basic type, following [7], this paper also considers HAR-RRV-RBV and HAR-RRV-RTV model. To eliminate the bias, the HAR-RRV-RBV model replays RRV_t by $\widehat{RRV}_t = \lambda(2)RRV_t + (1 - \lambda(2))RBV_t$, and gets

$$\widehat{RRV}_t = \beta + \beta_d RBV_{t-1} + \beta_w \overline{RBV}_{5,t-1} + \beta_m \overline{RBV}_{22,t-1} + \epsilon_t, \tag{7}$$

where $\overline{RBV}_{n,t}$ is the nth moving average at time t.

The HAR-RRV-RTV model replays RRV_t by $\widehat{RRV}_t = \lambda(2)RRV_t + (1 - \lambda(2))RTV_t$, and gets

$$\widehat{RRV}_t = \beta + \beta_d RTV_{t-1} + \beta_w \overline{RTV}_{5,t-1} + \beta_m \overline{RTV}_{22,t-1} + \epsilon_t, \tag{8}$$

where $\overline{RTV}_{n,t}$ is the nth moving average at time t.

Ripple and Moosa [15] argued that the futures markets differ from the stock markets in many respects and the OI provides additional information due to the complex relationship between open interest and trading volume. This paper adds the daily difference of logarithm of OI as a variable in HAR-RV type model for futures data. To research whether OI has influence on Chinese futures price, above models are chosen as benchmark. Then HAR-RV-OI models are defined by

$$\hat{\sigma}_t = \beta + \beta_d \hat{\sigma}_{t-1} + \beta_w \bar{\sigma}_{5,t-1} + \beta_m \bar{\sigma}_{22,t-1} + \beta_{od} \Delta OI_{t-1} + \epsilon_t, \tag{9}$$

$$\widehat{RRV}_t = \beta + \beta_d RBV_{t-1} + \beta_w \overline{RBV}_{5,t-1} + \beta_m \overline{RBV}_{22,t-1} + \beta_{od} \Delta OI_{t-1} + \epsilon_t, \tag{10}$$

$$\widehat{RRV}_t = \beta + \beta_d RTV_{t-1} + \beta_w \overline{RTV}_{5,t-1} + \beta_m \overline{RTV}_{22,t-1} + \beta_{od} \Delta OI_{t-1} + \epsilon_t, \tag{11}$$

respectively, where

$$\Delta OI_{t-1} = \log OI_{t-1} - \log OI_{t-2}. \tag{12}$$

Whether the basic HAR-RV model or the HAR-RV-OI model, the assumption is that there are linear relationships between the volatility estimator and its moving average or open interest. Besides the performance, this paper also considers the p-value of the models' coefficients. Since the HAR-RV type model employs moving average, it will trend to a horizontal line when forecasting.

2.3.2. Cyclical Two-Component Models

Let L_t denote the long trend component of the square root of volatility estimator, and S_t denote the short-run component. Then the square root of volatility estimator $\underline{\sigma}_t$ is the sum of L_t and S_t by the definition. Constructing a CTCM type model follows 3 steps.

First, calculate the long component, L_t. This paper uses low-pass filter of Hodrick and Prescott [16] to the price following [4]. Then we employ the filtered price in volatility estimator to calculate L_t. Harris et al. [4] employed filter in price rather than volatility estimator, due to the fact that the intraday prices are more likely to satisfy the assumptions of the non-parametric filters and thus provides reasonable estimations of the underlying long-run trends in volatility. This process is implemented by a MATLAB built-in function, *hpfilter()*.

Secondly, estimate an AR(1) model for $S_t = \underline{\sigma}_t - L_t$, i.e.,

$$S_t = \alpha S_{t-1} + \beta, \tag{13}$$

where β is a non-zero number. It differs from the assumption of [4], which set the constant term as a zero-mean random error. This change is based on the following reasons. The figures show that the short component is near zero, but its mean is non-zero. Considering the magnitude of volatility estimator is small, the result near zero cannot be seen as zero. On the other hand, in an AR(1) model, the constant term is usually a non-zero number. If the constant term is zero, that means the S_t will be zero when S_{t-1} is zero. On the other hand, the zero constant term means that today's short-run component is times or a percentage of the last day's short-run component. That does not satisfy the definition, which hints that the short-run component is a mean-reverting process.

Finally, given the $\hat{\alpha}$ estimated in the last step, the n-step ahead forecast of the square root of volatility estimator in CTCM is defined by

$$\underline{\sigma}_{F,t+n} = L_{F,t+n} + S_{F,t+n} = L_{F,n+t} + \hat{\alpha}^n(\underline{\sigma}_t - L_t) + \beta \cdot \frac{\hat{\alpha}^n - 1}{\hat{\alpha} - 1}, \tag{14}$$

where $L_{F,t+n} = L_t$ for any $n \geq 1$, for convenience. This follows [4], which claimed that the long-term component was a random walk.

This paper considers CTCM daily forecasting ($n = 1$), weekly forecasting ($n = 5$), monthly forecasting ($n = 22$), and totally 3 kinds of CTCM models.

As a comparison, we construct a CTCM type model with random walk term e_t.

$$\underline{\sigma}_{F,t+1} = L_t + E[S_t] + e_t, \tag{15}$$

where L_t is the long-term component calculated by the same way as above, $E[S_t]$ is the mean of short-term component, and e_t is white noise with zero-mean and $Var(e_t) = Var(S_t)$.

2.4. Estimation and Forecasting Methods

This paper employs in-sample estimation method to evaluate whether a model captures the characteristics of all data efficiently or not. Usually, effective estimation is essential to obtain quality forecasting results.

During forecasting, the data are divided into 2 groups. The first group is used to calculate the parameter of the model, and the second group is used to evaluate the performance of the model. We employ out of sample forecasting method with rolling windows.

2.5. Evaluation Rules

We employ Mincer–Zarnowitz regression test to evaluate the performance of estimation, and uses mean absolute percent error , root mean square error(RMSE), and Theil's U decomposition (Theil's U) to evaluate the forecasting results. In addition, we use modified Diebold–Mariano test (MDM-test) to diagnose whether the most efficient model has similar performance to other models.

2.5.1. Mincer–Zarnowitz Regression Test

Since CTCM is a non-linear regression model, goodness of fit R^2 is not a proper evaluation rule. Mincer–Zarnowitz regression test, which constructs a linear relationship between the estimation results and the original values, is a good replacement here.

Let $\hat{\sigma}_t$ denote the t-th day's volatility estimator calculated by the intraday data, and σ_{E_t} denote the estimation result for t-th day. Then the Mincer–Zarnowitz regression is defined as

$$\hat{\sigma}_t = \alpha_1 + \alpha_2 \sigma_{E_t} + \epsilon_t, \tag{16}$$

where α_1 and α_2 are the coefficients, and ϵ_t is the white noise with zero-mean. To evaluate the performance of model estimation, the value of goodness-of-fit R^2 of Mincer–Zarnowitz regression is used. The model with the highest MZ-R^2 is the best estimation for the products.

2.5.2. Root Mean Square Error

The RMSE is a method to describe the difference between forecasting value and real data. If a model has least value of RMSE, we can conclude that its value is nearest to the actual value. Then we can claim it is the most efficient model.

Let $\hat{\sigma}_t$ denote the volatility estimator calculated by the data and σ_{F_t} denotes the forecasting result. Then the RMSE is defined as

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\sigma_{F_t} - \hat{\sigma}_t)^2} \tag{17}$$

2.5.3. Mean Absolute Percent Error

When the volatility estimators are all small, a smallest RMSE cannot be evidence that a model has the most accurate forecasting results. Then other evaluations can be used in practice. MAPE is a method to describes the percentage difference between the forecasting results and the original value. Suppose the tolerance error is $\alpha\%$, which means only the model with the MAPE lower than $\alpha\%$ can be used in practice, and the forecasting results for this model is lower or higher $\alpha\%$ than the actual value, averagely.

Let $\hat{\sigma}_t$ denote the volatility estimator calculated by the data, and σ_{F_t} denote the forecasting result. Then the MAPE is defined as

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\frac{|\sigma_{F_t} - \hat{\sigma}_t|}{|\hat{\sigma}_t|}. \tag{18}$$

This paper follows that the model with the MAPE lower than 35% is feasible in practice, and the model with the lowest MAPE has the best forecast for those products.

2.5.4. Theil's U Decomposition

Theil's U statistic is a relatively accurate measure that compare the forecasted results with the results of forecasting with minimal historical data.

Let $\hat{\sigma}_t$ denote the volatility estimator calculated by the data, σ_{F_t} denote the forecasting result. Then the MAPE is defined as

$$U = \sqrt{\frac{\sum_{t=1}^{n-1}(\frac{\sigma_{F_{t+1}}-\hat{\sigma}_{t+1}}{\hat{\sigma}_t})^2}{\sum_{t=1}^{n-1}(\frac{\hat{\sigma}_{t+1}-\hat{\sigma}_t}{\hat{\sigma}_t})^2}} \tag{19}$$

If the U value is less than 1, the forecasting method is better than guessing.

2.5.5. Modified Diebold–Mariano Test

MDM-test calculates a measure of predictive accuracy. The hypothesis is that two model have the same forecasting performance.

Let e_{it} and e_{jt} be the forecasting error of model i and model j at time t, respectively. Define $d_t = e_{it}^2 - e_{jt}^2$ as the loss function. If model i and model j have the same forecasting behavior, $E[d_t] = 0$. Thus, the modified Diebold–Mariano test is defined as

$$\left[\frac{T+1-2h+T^{-1}h(h-1)}{T}\right]^{1/2} \bar{d}/V_h(\hat{\bar{d}})^{1/2}, \tag{20}$$

where $V_h(\hat{\bar{d}}) = T^{-1}[\gamma_0 + 2\sum_{k=1}^{h-1}\gamma_k]$, γ_k is kth-order covariance of sequence d_t, and h is the order of out of sample forecasting.

In our experiments, h is 1. If the p-value is less than -1.96 or greater than 1.96, we reject the hypothesis and conclude that the model has different forecasting performance.

3. Results

This paper aims to use cyclical two-component model (CTCM) with an appropriate volatility estimator to describe and forecast the Chinese futures markets as well as sector stocks. 3 kinds of CTCM type models are provided and tested. There are 21 CTCM type models (see Table 4) employed in experiment, which are: improved CTCM with RRV (denoted as CTCM-RRV-D, CTCM-RRV-W, CTCM-RRV-M, respectively), improved CTCM with RBV (denoted as CTCM-RBV-D, CTCM-RBV-W, CTCM-RBV-M, respectively), improved CTCM with RTV (denoting as CTCM-RTV-D, CTCM-RTV-W, CTCM-RTV-M, respectively), Harris' CTCM [4] with RRV (denoting as TWO-RRV-D, TWO-RRV-W, TWO-RRV-M, respectively), Harris' CTCM with RBV (denoting as TWO-RBV-D, TWO-RBV-W, TWO-RBV-M, respectively), Harris' CTCM with RTV (denoting as TWO-RTV-D, TWO-RTV-W, TWO-RTV-M, respectively), daily CTCM Random Walk model with RRV (denoting as TOWRW-RRV), daily CTCM Random Walk model with RBV (denoting as TWORW-RBV), daily CTCM Random Walk model with RTV (denoting as TWORW-RTV). As a benchmark, the performance of HAR-RV type models is compared with CTCM's. There are 10 HAR-RV type models used to be benchmark. They are: HAR-RRV, HAR-RRV-OI, HAR-RBV, HAR-RBV-OI, HAR-RTV, HAR-RTV-OI, HAR-RRV-RBV, HAR-RRV-RBV-OI, HAR-RRV-RTV, and HAR-RRV-RTV-OI. There are 12 products, covering 8 futures and 4 sector stocks.

The experiment has 3 parts. The first is using all the data to test the fitness of models. The results are represented by MZ-R^2, whose range is from 0 to 1. The higher value means the better estimation. The second is out of sample forecasting to check the forecasting fitness of models. The results are represented by MAPE, RMSE, and Theil's U statistics which have lower values mean they have better performance. The last is the summary of experiment results.

Table 4. The results of estimation.

	Ag	Al	Cu	I	J	JM	M	RM	ENG	MTR	MDC	FINRE
CTCM-RRV-D	0.5034	0.5095	0.3556	0.5055	0.5807	0.6913	0.5045	0.5223	0.6917	0.6571	0.6625	0.6879
CTCM-RRV-W	0.1910	0.1930	0.0978	0.2486	0.2874	0.3982	0.2218	0.2444	0.2558	0.2524	0.3218	0.2485
CTCM-RRV-M	0.0221	0.0604	0.0085	0.0826	0.0845	0.1532	0.1123	0.1081	0.2412	0.3304	0.3404	0.1081
CTCM-RBV-D	0.5763	0.5548	0.5189	0.5336	0.6032	0.7026	0.5717	0.5618	0.7494	0.7104	0.7205	0.7469
CTCM-RBV-W	0.2134	0.2089	0.1399	0.2762	0.3018	0.4037	0.2422	0.2615	0.2828	0.2857	0.3567	0.2758
CTCM-RBV-M	0.0263	0.0661	0.0180	0.0901	0.0850	0.1524	0.1160	0.1160	0.1961	0.3007	0.2942	0.0804
CTCM-RTV-D	**0.6166**	**0.5750**	**0.5520**	0.5438	**0.6201**	**0.7181**	**0.6016**	**0.5765**	**0.7562**	**0.7121**	**0.7375**	**0.7532**
CTCM-RTV-W	0.2296	0.2188	0.1484	0.2806	0.3109	0.4082	0.2580	0.2708	0.2897	0.2918	0.3756	0.2840
CTCM-RTV-M	0.0296	0.0710	0.0202	0.0937	0.0871	0.1542	0.1230	0.1195	0.1871	0.2886	0.2723	0.0817
TWO-RRV-D	0.4667	0.4161	0.2950	0.3854	0.4816	0.5809	0.3991	0.4187	0.6524	0.6168	0.6123	0.6405
TWO-RRV-W	0.1915	0.2109	0.1006	0.2657	0.3193	0.4343	0.2332	0.2574	0.2534	0.2495	0.3171	0.2454
TWO-RRV-M	0.0198	0.0571	0.0089	0.0723	0.0790	0.1456	0.1036	0.0986	0.2526	0.3420	0.3479	0.1159
TWO-RBV-D	0.5556	0.4895	0.4522	0.4359	0.5244	0.6149	0.4729	0.4750	0.7487	0.7094	0.7160	0.7456
TWO-RBV-W	0.2122	0.2137	0.1412	0.2920	0.3287	0.4387	0.2501	0.2713	0.2825	0.2853	0.3555	0.2754
TWO-RBV-M	0.0251	0.0639	0.0180	0.0835	0.0817	0.1468	0.1103	0.1104	0.1974	0.3026	0.2969	0.0812
TWO-RTV-D	0.6065	0.5345	0.5136	0.4651	0.5556	0.6440	0.5224	0.5122	0.7539	0.7098	0.7368	0.7512
TWO-RTV-W	0.2283	0.2168	0.1456	0.2955	0.3314	0.4374	0.2623	0.2752	0.2902	0.2923	0.3760	0.2846
TWO-RTV-M	0.0288	0.0692	0.0201	0.0885	0.0844	0.1494	0.1185	0.1154	0.1850	0.2859	0.2713	0.0808
TWORW-RRV	0.3852	0.4434	0.3160	0.3644	0.3963	0.5680	0.4190	0.3448	0.5753	0.5488	0.5079	0.5705
TWORW-RBV	0.5392	0.4956	0.4708	0.4535	0.4532	0.6137	0.4673	0.4858	0.6670	0.5989	0.6205	0.6278
TWORW-RTV	0.5703	0.5207	0.5063	0.4992	0.5160	0.6581	0.5285	0.5299	0.6742	0.6148	0.6471	0.6403
HAR-RRV	0.4493	0.4295	0.3128	0.5105	0.4834	0.6091	0.3835	0.4145	0.5682	0.5110	0.5154	0.4743
HAR-RBV	0.5206	0.4634	0.4436	0.5766	0.5378	0.6466	0.4683	0.4808	0.5514	0.5263	0.5355	0.5062
HAR-RTV	0.5543	0.4702	0.4596	0.5778	0.5602	0.6593	0.5029	0.4946	0.5411	0.5070	0.5385	0.5120
HAR-RRV-OI	0.4518	0.4304	0.3129	0.5145	0.4834	0.6091	0.3851	0.4145				
HAR-RBV-OI	0.5208	0.4643	0.4436	0.5795	0.5380	0.6466	0.4700	0.4808				
HAR-RTV-OI	0.5543	0.4712	0.4596	**0.5806**	0.5605	0.6593	0.5042	0.4946				
HRA-RRV-RBV	0.3666	0.3684	0.1728	0.4119	0.4238	0.5825	0.2990	0.3393	0.4510	0.4226	0.4222	0.3959
HAR-RRV-RTV	0.3667	0.3422	0.1673	0.4031	0.4065	0.5642	0.2831	0.3273	0.4188	0.3928	0.3883	0.3797
HAR-RRV-RBV-OI	0.3674	0.3687	0.1731	0.4178	0.4242	0.5826	0.3010	0.3393				
HAR-RRV-RTV-OI	0.3671	0.3425	0.1677	0.4097	0.4071	0.5642	0.2855	0.3273				

3.1. Estimation Results

The numerical results of estimation are listed in the following table (Table 4). The results are between 0 and 1, and the model with highest value is the most efficient one for the product. Besides the product I, most products are best estimated by the improved daily cyclical two-component model with realized range-based tripower volatility (CTCM-RTV-D).

For space saving, figures of the best estimation for some products are shown below (Figures 1–3). The blue line is the actual value, while the orange one is the estimated value. We can find that all the estimation lines are near the actual line.

3.2. Out of Sample Forecasting Results

The non-negative results of out of sample forecast are listed in the following tables (Table 5). The first table is based on the MAPE evaluation rule. Notice that the values in the following table are percentage. The model with the lowest value is the most efficient one for the product. Based on MAPE, Ag, Cu, ENG, FINRE, MDC, and MTR are best forecasted by the improved daily cyclical two-component model with realized range-based tripower volatility (CTCM-RTV-D). However, Al, Fe, J, JM, M, and RM are best forecasted by the Harris' daily cyclical two-component model with realized range-based tripower volatility (TWO-RTV-D).

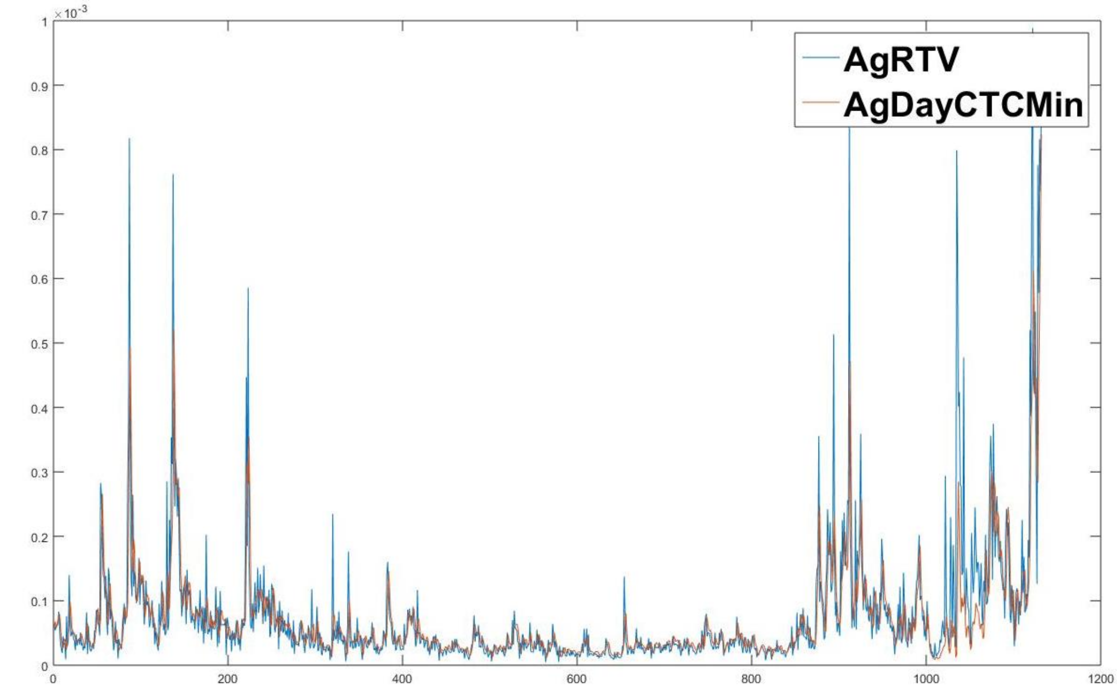

Figure 1. Model CTCM-RTV-D Estimation for Ag(SHFE). The horizontal line is the timeline, and the perpendicular line is the volatility estimator. The actual values are represented by the blue line, while the estimated values are represented by the orange line. During the Day 400 and the Day 800, when the volatility is smooth, the estimated values are almost consistent with the actual values. The CTCM-RTV-D model also captures the characteristics of the volatility when it is volatile.

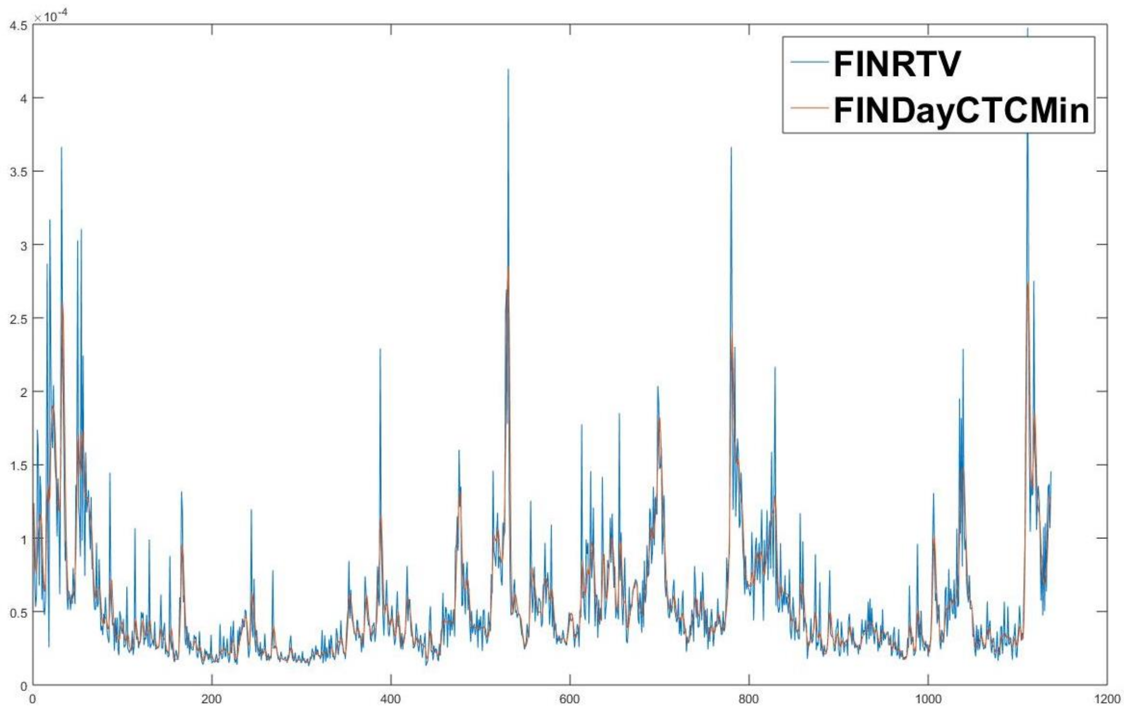

Figure 2. Model CTCM-RTV-D Estimation for FINRE. The horizontal line is the timeline, and the perpendicular line is the volatility estimator. The actual values are represented by the blue line, while the estimated values are represented by the orange line. We find that the estimated values are close to the actual values, although the volatility is high throughout the whole period.

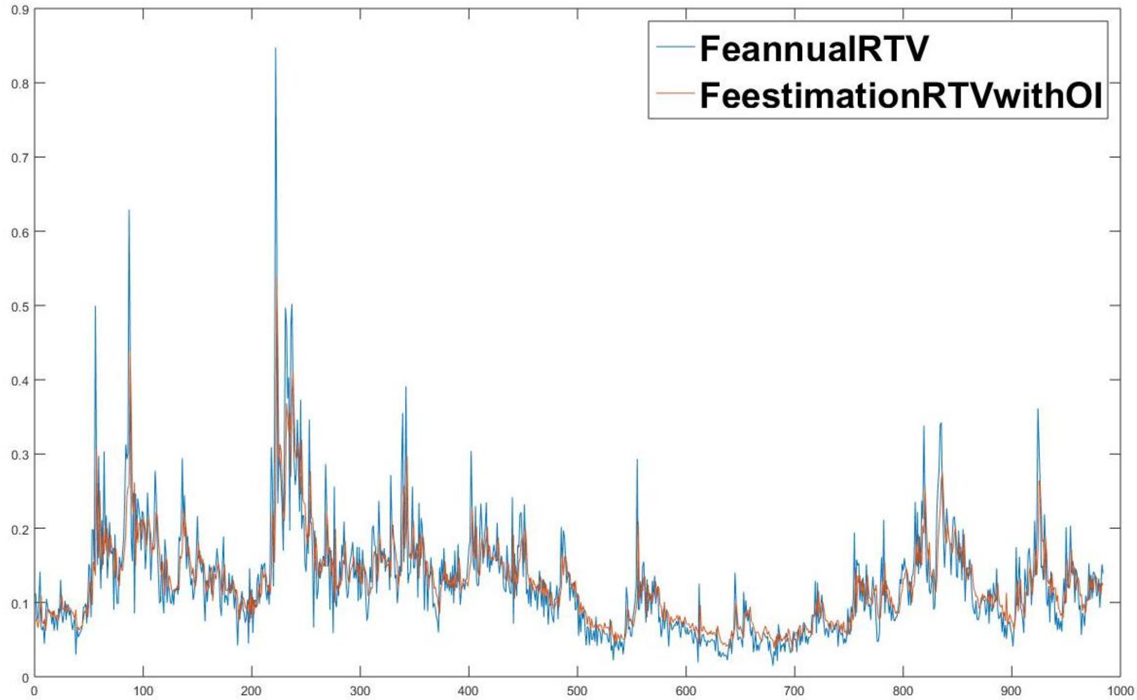

Figure 3. Model HAR-RTV-OI Estimation for I(DCE). The horizontal line is the timeline, and the perpendicular line is the volatility estimator. The actual values are represented by the blue line, while the estimated values are represented by the orange line. The estimated values are close to the actual values, and the trend of volatility is smooth, especially after Day 300.

Table 5. The results of out of sample forecast MAPE(%).

	Ag	Al	Cu	I	J	JM	M	RM	ENG	MTR	MDC	FINRE
CTCM-RRV-D	42.611	37.118	36.896	37.157	52.675	75.013	42.325	37.479	18.452	27.870	23.336	22.944
CTCM-RRV-W	55.424	50.458	45.927	45.130	60.902	86.047	59.738	47.992	30.167	42.531	36.054	38.187
CTCM-RRV-M	78.429	65.030	63.463	55.813	78.117	108.360	58.274	55.912	55.637	78.366	54.552	76.276
CTCM-RBV-D	40.601	33.249	34.663	36.300	52.131	65.151	40.468	36.380	13.608	23.962	20.288	17.906
CTCM-RBV-W	58.794	49.793	48.622	45.749	61.200	76.912	60.355	48.297	27.909	41.301	36.250	35.076
CTCM-RBV-M	82.288	67.086	68.418	56.514	81.740	102.061	57.122	56.580	52.265	80.512	59.894	74.046
CTCM-RTV-D	**39.551**	31.975	**34.142**	35.846	51.528	62.371	40.532	37.046	**12.610**	**22.613**	**18.899**	**16.694**
CTCM-RTV-W	60.519	49.687	51.307	45.982	61.039	75.099	61.460	49.563	27.295	41.180	36.745	34.568
CTCM-RTV-M	84.146	68.991	71.135	56.716	82.881	101.152	57.538	57.877	50.716	80.686	63.876	72.891
TWO-RRV-D	52.334	43.885	43.398	30.888	34.064	36.034	34.674	34.085	17.179	27.310	33.893	22.400
TWO-RRV-W	58.568	46.670	44.945	34.009	35.854	40.231	43.444	37.173	25.006	38.034	38.873	31.701
TWO-RRV-M	76.951	57.535	61.493	38.685	44.478	49.981	41.816	40.537	46.754	68.464	53.255	63.425
TWO-RBV-D	44.635	35.085	36.039	25.243	26.138	29.361	31.183	26.876	13.088	22.955	21.392	17.179
TWO-RBV-W	59.039	45.867	42.634	30.794	33.205	36.006	44.153	34.095	27.062	40.145	35.218	33.866
TWO-RBV-M	80.792	59.026	63.745	38.341	48.804	54.229	42.039	40.730	51.278	78.614	58.313	72.019
TWO-RTV-D	41.241	**31.971**	34.283	**23.772**	**25.707**	**27.429**	**30.700**	**26.091**	13.735	24.161	19.294	18.350
TWO-RTV-W	60.118	46.659	45.337	31.329	33.773	36.366	46.313	36.089	28.728	42.829	37.694	36.694
TWO-RTV-M	83.300	61.257	66.801	40.441	52.412	58.784	43.901	43.141	52.511	83.153	64.795	76.251
TWORW-RRV	48.449	51.589	46.472	40.854	61.856	94.345	59.848	45.150	32.954	39.390	35.181	31.552
TWORW-RBV	45.608	44.364	44.858	39.802	59.905	82.709	42.998	40.597	26.663	37.195	31.732	34.054
TWORW-RTV	44.373	42.951	41.727	42.107	63.824	75.523	44.505	38.382	23.621	35.229	30.228	30.307
HAR-RRV	51.652	51.840	65.125	28.763	43.740	44.772	40.432	32.969	23.352	41.176	33.403	32.964
HAR-RBV	49.818	48.391	50.490	25.188	38.635	40.434	36.717	28.603	20.943	36.257	29.415	28.959
HAR-RTV	48.261	46.940	50.407	24.755	36.800	39.582	36.251	28.778	19.887	34.464	27.594	27.655
HAR-RRV-OI	50.684	52.740	65.369	29.968	44.401	44.593	41.586	32.964				
HAR-RBV-OI	50.073	49.229	50.853	26.717	38.698	40.325	37.775	28.586				
HAR-RTV-OI	48.671	48.349	50.793	26.192	36.525	39.453	37.088	28.765				
HRA-RRV-RBV	48.558	47.817	45.511	30.135	42.792	50.593	44.795	35.178	39.400	46.742	38.448	44.660
HAR-RRV-RTV	48.037	48.009	44.392	30.788	42.461	52.919	45.405	35.288	46.497	47.920	38.631	48.097
HAR-RRV-RBV-OI	48.320	48.898	45.642	31.811	44.304	50.273	46.434	35.355				
HAR-RRV-RTV-OI	48.018	49.015	44.303	32.621	44.282	52.575	47.345	35.606				

By the evaluation rule of this paper, the models can be employed in practice if and only if the MAPEs are less than 35%. Thus, all the most efficient models above satisfy the rule, except for Ag(SHFE).

For space saving, figures of the best forecast for some products are shown below (Figures 4–6). The blue line is the actual value, while the orange one is the forecasted value.

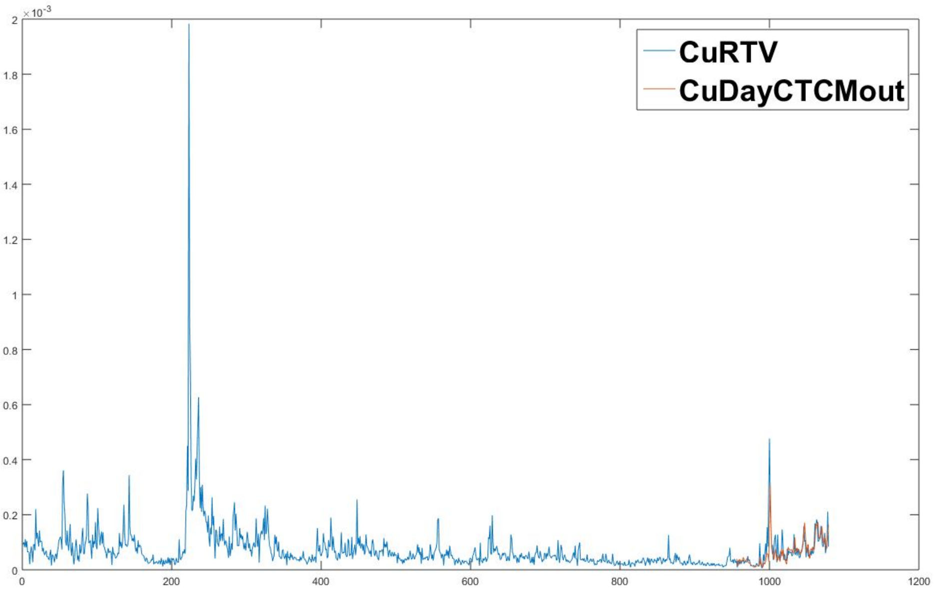

Figure 4. Model CTCM-RTV-D Forecast for Cu(SHFE). The horizontal line is the timeline, and the perpendicular line is the volatility estimator. The forecasted period is the last 124 days. The actual values are represented by the blue line, while the forecasted values are represented by the orange line. The forecasted values are closed to the actual values, although there is a sharp jump around the Day 1000.

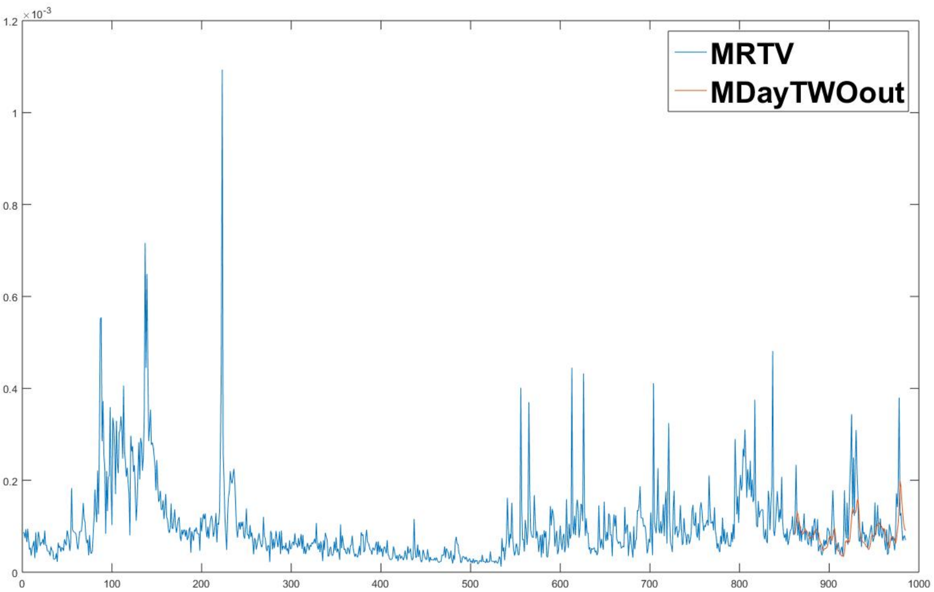

Figure 5. Model TWO-RTV-D Forecast for M(DCE). The horizontal line is the timeline, and the perpendicular line is the volatility estimator. The forecasted period is the last 124 days. The actual values are represented by the blue line, while the forecasted values are represented by the orange line. The forecasted values are closed to the actual values, except for a sharp jump around the Day 920.

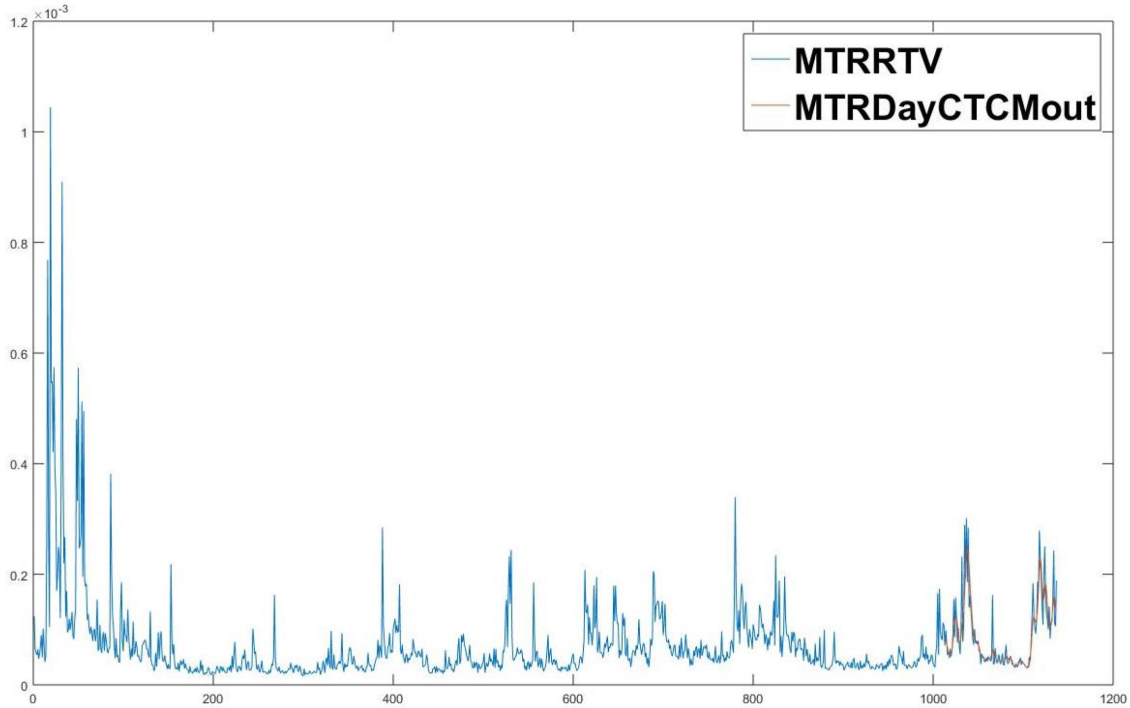

Figure 6. Model CTCM-RTV-D Forecast for MTR. The horizontal line is the timeline, and the perpendicular line is the volatility estimator. The forecasted period is the last 124 days. The actual values are represented by the blue line, while the forecasted values are represented by the orange line. The forecasted values are closed to the actual values, except for a sharp jump around the Day 1100.

The second table is based on the RMSE evaluation rule (Table 6). The model with the lowest value is the most efficient one for the product. Based on RMSE, Ag, Al, Cu, Fe, and M are best forecasted by the improved daily cyclical two-component model with realized range-based tripower volatility (CTCM-RTV-D). However, J, JM, ENG, FIN, MDC and MTR are best forecasted by the Harris' daily cyclical two-component model with realized range-based tripower volatility (TWO-RTV-D). RM is best forecasted by cyclical two-component model with random walk term with realized range-based tripower volatility (TWORW-RTV).

For space saving, figures of the best forecast for some products are shown below (Figures 7–9). The blue line is the actual value, while the orange one is the forecasted value.

Table 6. The results of out of sample forecast RMSE.

	Ag	**Al**	**Cu**	**I**	**J**	**JM**	**M**	**RM**	**ENG**	**MTR**	**MDC**	**FINRE**
CTCM-RRV-D	2.3×10^{-4}	9.9×10^{-5}	2.0×10^{-4}	2.3×10^{-4}	1.6×10^{-4}	1.4×10^{-4}	6.7×10^{-5}	1.1×10^{-4}	3.7×10^{-5}	5.3×10^{-5}	6.9×10^{-5}	5.6×10^{-5}
CTCM-RRV-W	2.9×10^{-4}	1.1×10^{-4}	2.2×10^{-4}	2.8×10^{-4}	1.8×10^{-4}	1.6×10^{-4}	9.1×10^{-5}	1.4×10^{-4}	5.9×10^{-5}	8.6×10^{-5}	9.8×10^{-5}	8.0×10^{-5}
CTCM-RRV-M	3.6×10^{-4}	1.3×10^{-4}	2.4×10^{-4}	3.5×10^{-4}	2.2×10^{-4}	2.1×10^{-4}	9.4×10^{-5}	1.6×10^{-4}	8.0×10^{-5}	1.1×10^{-4}	1.4×10^{-4}	1.1×10^{-4}
CTCM-RBV-D	1.8×10^{-4}	5.3×10^{-5}	5.8×10^{-5}	1.9×10^{-4}	1.3×10^{-4}	1.1×10^{-4}	5.3×10^{-5}	7.9×10^{-5}	2.7×10^{-5}	3.5×10^{-5}	4.4×10^{-5}	3.5×10^{-5}
CTCM-RBV-W	2.4×10^{-4}	6.9×10^{-5}	7.9×10^{-5}	2.5×10^{-4}	1.6×10^{-4}	1.3×10^{-4}	7.8×10^{-5}	1.1×10^{-4}	5.1×10^{-5}	6.9×10^{-5}	7.8×10^{-5}	6.3×10^{-5}
CTCM-RBV-M	3.0×10^{-4}	8.0×10^{-5}	9.4×10^{-5}	3.1×10^{-4}	1.9×10^{-4}	1.7×10^{-4}	7.8×10^{-5}	1.3×10^{-4}	7.2×10^{-5}	9.4×10^{-5}	1.2×10^{-4}	9.4×10^{-5}
CTCM-RTV-D	$\mathbf{1.4 \times 10^{-4}}$	$\mathbf{4.0 \times 10^{-5}}$	$\mathbf{4.6 \times 10^{-5}}$	$\mathbf{1.8 \times 10^{-4}}$	1.2×10^{-4}	9.6×10^{-5}	$\mathbf{4.9 \times 10^{-5}}$	7.5×10^{-5}	2.3×10^{-5}	2.9×10^{-5}	3.7×10^{-5}	3.0×10^{-5}
CTCM-RTV-W	2.0×10^{-4}	5.7×10^{-5}	6.8×10^{-5}	2.4×10^{-4}	1.4×10^{-4}	1.2×10^{-4}	7.3×10^{-5}	1.0×10^{-4}	4.7×10^{-5}	6.4×10^{-5}	7.0×10^{-5}	5.8×10^{-5}
CTCM-RTV-M	2.6×10^{-4}	6.7×10^{-5}	8.2×10^{-5}	3.0×10^{-4}	1.8×10^{-4}	1.6×10^{-4}	7.3×10^{-5}	1.2×10^{-4}	6.8×10^{-5}	8.8×10^{-5}	1.1×10^{-4}	9.0×10^{-5}
TWO-RRV-D	2.6×10^{-4}	1.1×10^{-4}	2.2×10^{-4}	3.1×10^{-4}	1.9×10^{-4}	1.4×10^{-4}	8.8×10^{-5}	1.5×10^{-4}	4.4×10^{-5}	6.3×10^{-5}	8.7×10^{-5}	6.5×10^{-5}
TWO-RRV-W	3.0×10^{-4}	1.2×10^{-4}	2.2×10^{-4}	3.3×10^{-4}	2.0×10^{-4}	1.5×10^{-4}	9.7×10^{-5}	1.6×10^{-4}	6.1×10^{-5}	8.8×10^{-5}	1.1×10^{-4}	8.5×10^{-5}
TWO-RRV-M	3.7×10^{-4}	1.4×10^{-4}	2.5×10^{-4}	3.8×10^{-4}	2.2×10^{-4}	1.9×10^{-4}	1.0×10^{-4}	1.8×10^{-4}	8.1×10^{-5}	1.1×10^{-4}	1.5×10^{-4}	1.1×10^{-4}
TWO-RBV-D	1.8×10^{-4}	6.2×10^{-5}	6.8×10^{-5}	2.2×10^{-4}	1.3×10^{-4}	9.1×10^{-5}	6.2×10^{-5}	9.4×10^{-5}	2.8×10^{-5}	3.5×10^{-5}	4.6×10^{-5}	3.6×10^{-5}
TWO-RBV-W	2.4×10^{-4}	7.2×10^{-5}	8.1×10^{-5}	2.5×10^{-4}	1.4×10^{-4}	1.0×10^{-4}	7.5×10^{-5}	1.1×10^{-4}	5.1×10^{-5}	6.9×10^{-5}	7.9×10^{-5}	6.3×10^{-5}
TWO-RBV-M	3.1×10^{-4}	8.5×10^{-5}	9.7×10^{-5}	3.1×10^{-4}	1.7×10^{-4}	1.4×10^{-4}	8.0×10^{-5}	1.3×10^{-4}	7.2×10^{-5}	9.4×10^{-5}	1.2×10^{-4}	9.4×10^{-5}
TWO-RTV-D	1.4×10^{-4}	4.5×10^{-5}	5.2×10^{-5}	1.8×10^{-4}	$\mathbf{1.1 \times 10^{-4}}$	$\mathbf{7.5 \times 10^{-5}}$	5.4×10^{-5}	8.0×10^{-5}	$\mathbf{2.3 \times 10^{-5}}$	$\mathbf{2.9 \times 10^{-5}}$	$\mathbf{3.6 \times 10^{-5}}$	$\mathbf{3.0 \times 10^{-5}}$
TWO-RTV-W	2.0×10^{-4}	5.8×10^{-5}	6.9×10^{-5}	2.2×10^{-4}	1.2×10^{-4}	9.1×10^{-5}	6.9×10^{-5}	1.0×10^{-4}	4.7×10^{-5}	6.4×10^{-5}	7.0×10^{-5}	5.8×10^{-5}
TWO-RTV-M	2.7×10^{-4}	7.1×10^{-5}	8.3×10^{-5}	2.8×10^{-4}	1.5×10^{-4}	1.3×10^{-4}	7.2×10^{-5}	1.2×10^{-4}	6.8×10^{-5}	8.8×10^{-5}	1.1×10^{-4}	9.0×10^{-5}
TWORW-RRV	2.3×10^{-4}	1.0×10^{-4}	2.1×10^{-4}	2.7×10^{-4}	2.0×10^{-4}	1.9×10^{-4}	8.0×10^{-5}	1.3×10^{-4}	4.6×10^{-5}	5.3×10^{-5}	7.8×10^{-5}	5.9×10^{-5}
TWORW-RBV	1.8×10^{-4}	5.7×10^{-5}	6.2×10^{-5}	2.3×10^{-4}	1.6×10^{-4}	1.5×10^{-4}	5.8×10^{-5}	8.5×10^{-5}	3.6×10^{-5}	3.7×10^{-5}	5.3×10^{-5}	4.3×10^{-5}
TWORW-RTV	1.5×10^{-4}	4.7×10^{-5}	5.2×10^{-5}	2.1×10^{-4}	1.5×10^{-4}	1.3×10^{-4}	5.7×10^{-5}	$\mathbf{7.3 \times 10^{-5}}$	2.8×10^{-5}	3.4×10^{-5}	4.6×10^{-5}	4.1×10^{-5}
HAR-RRV	5.6×10^{-2}	2.5×10^{-2}	5.1×10^{-2}	5.2×10^{-2}	3.6×10^{-2}	2.5×10^{-2}	1.7×10^{-2}	2.9×10^{-2}	1.1×10^{-2}	1.6×10^{-2}	2.0×10^{-2}	1.6×10^{-2}
HAR-RBV	4.5×10^{-2}	1.4×10^{-2}	1.3×10^{-2}	4.0×10^{-2}	2.8×10^{-2}	1.8×10^{-2}	1.4×10^{-2}	2.1×10^{-2}	9.5×10^{-3}	1.3×10^{-2}	1.6×10^{-2}	1.2×10^{-2}
HAR-RTV	3.7×10^{-2}	1.1×10^{-2}	1.2×10^{-2}	3.7×10^{-2}	2.4×10^{-2}	1.7×10^{-2}	1.3×10^{-2}	2.0×10^{-2}	8.6×10^{-3}	1.1×10^{-2}	1.4×10^{-2}	1.1×10^{-2}
HAR-RRV-OI	5.6×10^{-2}	2.5×10^{-2}	5.1×10^{-2}	5.3×10^{-2}	3.7×10^{-2}	2.5×10^{-2}	1.8×10^{-2}	2.9×10^{-2}				
HAR-RBV-OI	4.5×10^{-2}	1.4×10^{-2}	1.4×10^{-2}	4.1×10^{-2}	2.8×10^{-2}	1.8×10^{-2}	1.4×10^{-2}	2.1×10^{-2}				
HAR-RTV-OI	3.7×10^{-2}	1.1×10^{-2}	1.2×10^{-2}	3.8×10^{-2}	2.4×10^{-2}	1.7×10^{-2}	1.3×10^{-2}	2.0×10^{-2}				
HRA-RRV-RBV	7.5×10^{-2}	4.5×10^{-2}	1.2×10^{-1}	8.7×10^{-2}	5.2×10^{-2}	4.2×10^{-2}	2.6×10^{-2}	4.6×10^{-2}	1.6×10^{-2}	2.5×10^{-2}	2.9×10^{-2}	2.5×10^{-2}
HAR-RRV-RTV	9.1×10^{-2}	5.2×10^{-2}	1.3×10^{-1}	9.4×10^{-2}	6.0×10^{-2}	4.7×10^{-2}	2.8×10^{-2}	4.9×10^{-2}	1.8×10^{-2}	2.8×10^{-2}	3.3×10^{-2}	2.7×10^{-2}
HAR-RRV-RBV-OI	7.4×10^{-2}	4.5×10^{-2}	1.2×10^{-1}	8.7×10^{-2}	5.2×10^{-2}	4.2×10^{-2}	2.6×10^{-2}	4.6×10^{-2}				
HAR-RRV-RTV-OI	9.1×10^{-2}	5.2×10^{-2}	1.3×10^{-1}	9.4×10^{-2}	6.0×10^{-2}	4.7×10^{-2}	2.9×10^{-2}	4.9×10^{-2}				

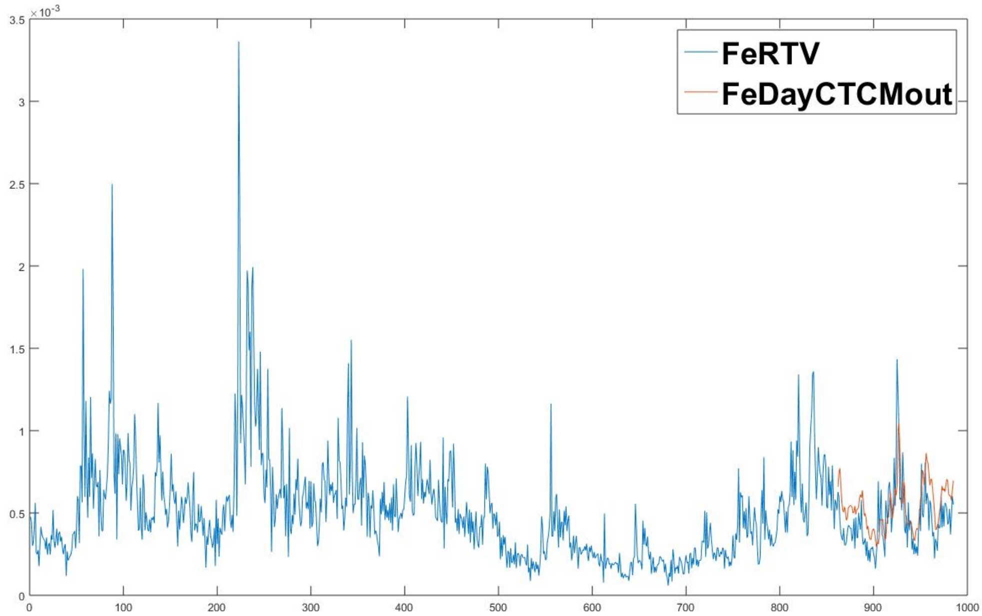

Figure 7. Model CTCM-RTV-D Forecast for I(DCE). The horizontal line is the timeline, and the perpendicular line is the volatility estimator. The forecasted period is the last 124 days. The actual values are represented by the blue line, while the forecasted values are represented by the orange line. The forecasted values are a bit higher than the actual values in most time.

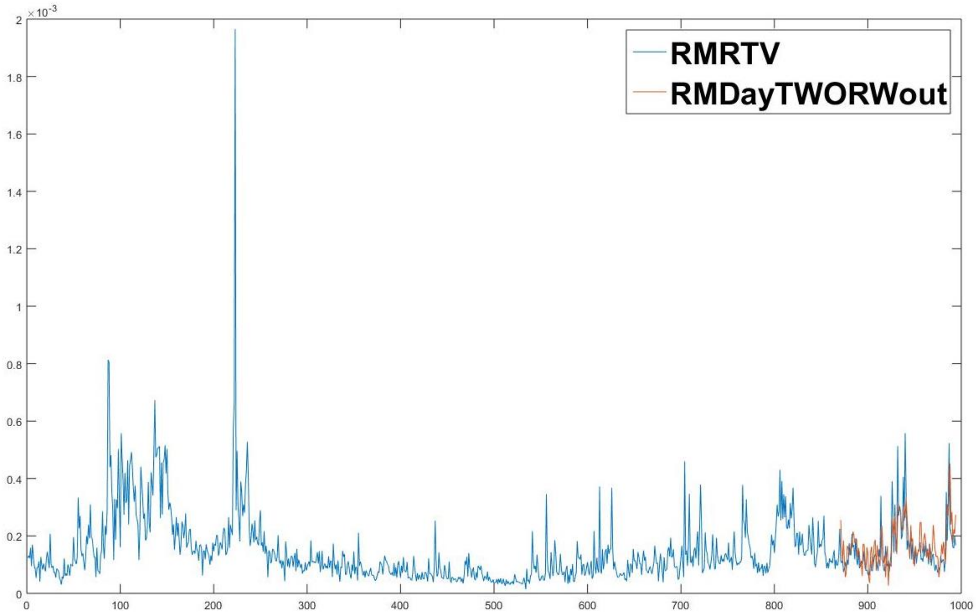

Figure 8. Model TWORW-RTV Forecast for RM(CZCE). The horizontal line is the timeline, and the perpendicular line is the volatility estimator. The forecasted period is the last 124 days. The actual values are represented by the blue line, while the forecasted values are represented by the orange line. The forecasted values are more volatile than the actual values, since the forecasted model is constructed by a random walk.

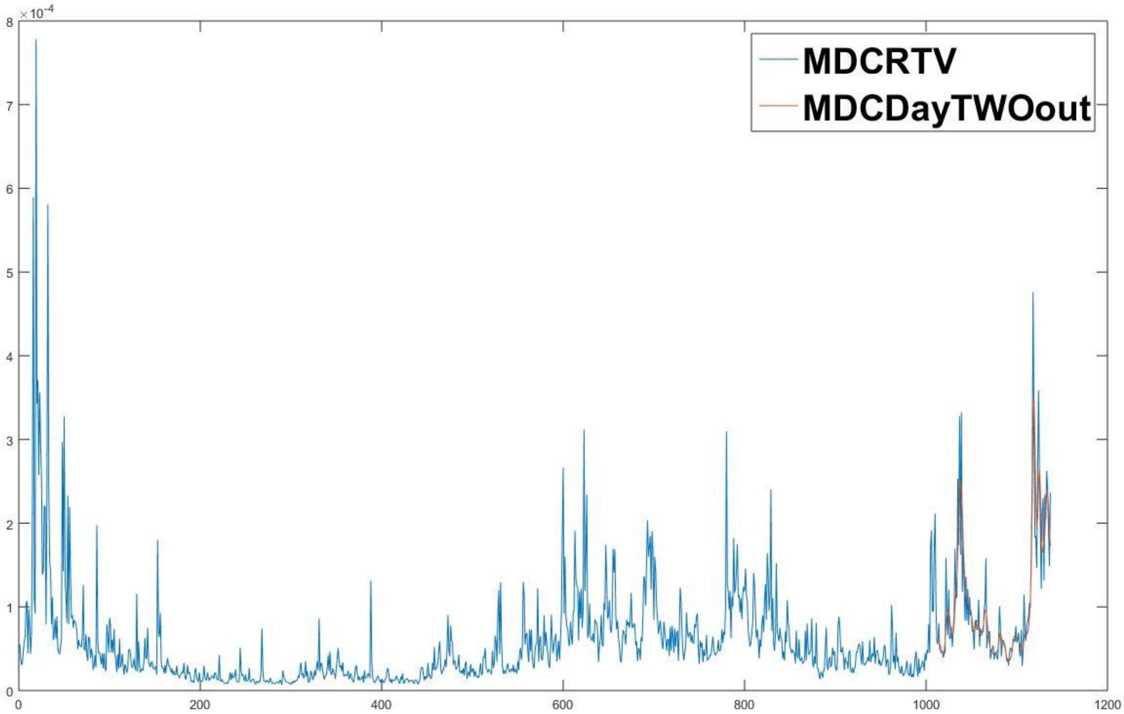

Figure 9. Model TWO-RTV-D Forecast for MDC. The horizontal line is the timeline, and the perpendicular line is the volatility estimator. The forecasted period is the last 124 days. The actual values are represented by the blue line, while the forecasted values are represented by the orange line. The forecasted values are close to the actual values, although there is a sharp jump around the Day 1100.

The third table (Table 7) is based on the Theil's U evaluation rule statistics. The model with the lowest value is the most efficient one for the product. Based on Theil's U statistics, Ag, Cu, Fe, J, and JM are best forecasted by HAR-RV type model. Al, and M are best forecasted by the improved daily cyclical two-component model with realized range-based tripower volatility (CTCM-RTV-D). However, ENG, MTR, MDC and FINRE are best forecasted by the Harris' daily cyclical two-component model with realized range-based tripower volatility (TWO-RTV-D).

For space saving, figures of the best forecast for some products are shown below (Figures 10–12). The blue line is the actual value, while the orange one is the forecast value.

Table 7. The results of out of sample forecast Theil's U.

	Ag	Al	Cu	I	J	JM	M	RM	ENG	MTR	MDC	FINRE
CTCM-RRV-D	1.0383	0.9644	0.9866	1.0278	0.9293	1.3912	0.9178	**0.9111**	0.7801	0.7921	0.8451	0.8369
CTCM-RRV-W	1.0230	1.0388	0.9920	1.2798	1.1120	1.7633	1.4224	1.1515	1.1719	1.0197	1.0575	1.1499
CTCM-RRV-M	1.1173	1.0653	1.0061	1.8050	1.4690	2.2495	1.2788	1.2731	2.2661	2.0346	1.5510	2.7378
CTCM-RBV-D	1.0065	0.9348	0.9060	1.1391	1.0356	1.4573	0.9136	0.9330	0.6847	0.6816	0.7309	0.7447
CTCM-RBV-W	0.9965	1.0890	0.9673	1.5315	1.2923	1.9976	1.6910	1.3147	1.2545	1.0708	1.0864	1.2144
CTCM-RBV-M	1.0735	1.1919	1.2515	1.9695	1.7536	2.7428	1.4756	1.5065	2.3416	2.3369	1.7247	2.9496
CTCM-RTV-D	0.9977	**0.8893**	0.8936	1.1233	1.0725	1.3617	**0.8956**	0.9198	0.6675	0.6523	0.6681	0.7188
CTCM-RTV-W	0.9945	1.1154	1.0619	1.5546	1.3953	1.9517	1.7589	1.3437	1.2944	1.1154	1.1467	1.2412
CTCM-RTV-M	1.0753	1.2440	1.4152	1.9995	1.9257	2.7181	1.5061	1.5417	2.3890	2.4932	2.0165	3.0257
TWO-RRV-D	1.0822	1.1130	1.0094	1.2233	1.1365	1.1665	1.1639	1.1714	0.9339	0.9442	1.0961	0.9935
TWO-RRV-W	1.0578	1.1251	1.0096	1.2478	1.1365	1.2057	1.2387	1.1903	1.1848	1.0710	1.1952	1.1808
TWO-RRV-M	1.1236	1.1151	1.0189	1.4378	1.1778	1.3050	1.1903	1.2257	2.0016	1.7986	1.5732	2.3850
TWO-RBV-D	1.0294	1.0892	1.0493	1.1246	1.0204	1.0464	1.0558	1.0625	0.6960	0.7024	0.7858	0.7593
TWO-RBV-W	1.0081	1.1407	1.0314	1.2296	1.0552	1.1886	1.3291	1.1559	1.2483	1.0704	1.1023	1.2099
TWO-RBV-M	1.0753	1.1716	1.2243	1.4924	1.2405	1.5442	1.2210	1.2866	2.3075	2.2892	1.6979	2.8975
TWO-RTV-D	1.0107	1.0398	1.0128	1.0434	0.9605	0.9604	0.9907	0.9865	**0.6562**	**0.6352**	**0.6524**	**0.7068**
TWO-RTV-W	0.9999	1.1436	1.0773	1.2072	1.0468	1.1745	1.4085	1.1582	1.3115	1.1230	1.1490	1.2552
TWO-RTV-M	1.0757	1.1974	1.3490	1.5322	1.3465	1.6374	1.2387	1.3099	2.4567	2.5609	2.0447	3.1148
TWORW-RRV	1.0408	1.0060	0.9884	1.1817	1.2810	1.9023	1.2592	1.1337	1.1834	0.8958	0.9608	1.0207
TWORW-RBV	1.0220	1.0536	0.9860	1.3543	1.2484	2.2352	1.0530	1.1446	1.0351	0.9599	0.9634	1.0584
TWORW-RTV	0.9984	1.0716	0.9802	1.4068	1.4783	1.8851	1.0989	1.0178	0.9803	0.8971	0.9462	1.0632
HAR-RRV	**0.9521**	0.9510	0.9657	0.9229	0.8481	0.9460	0.9193	0.9458	0.9567	0.9082	0.9272	0.9728
HAR-RBV	0.9555	0.9548	**0.8704**	0.9093	0.8767	0.9495	0.9137	0.9427	0.9754	0.9066	0.9254	0.9804
HAR-RTV	0.9590	0.9522	0.8938	**0.9047**	0.8859	0.9488	0.9159	0.9388	0.9821	0.9117	0.9471	0.9891
HAR-RRV-OI	0.9521	0.9570	0.9658	0.9283	0.8593	**0.9440**	0.9962	0.9458				
HAR-RBV-OI	0.9555	0.9953	0.8784	0.9493	0.8779	0.9486	1.0016	0.9429				
HAR-RTV-OI	0.9589	1.0290	0.9156	0.9527	0.8804	0.9506	0.9775	0.9389				
HRA-RRV-RBV	0.9639	0.9689	0.9889	0.9411	**0.8268**	0.9581	0.9249	0.9555	0.9803	0.9281	0.9255	0.9747
HAR-RRV-RTV	0.9733	0.9800	0.9899	0.9409	0.8370	0.9683	0.9278	0.9555	1.0255	0.9271	0.9233	0.9919
HAR-RRV-RBV-OI	0.9639	0.9672	0.9889	0.9322	0.8326	0.9555	0.9993	0.9543				
HAR-RRV-RTV-OI	0.9733	0.9785	0.9899	0.9319	0.8439	0.9656	1.0140	0.9537				

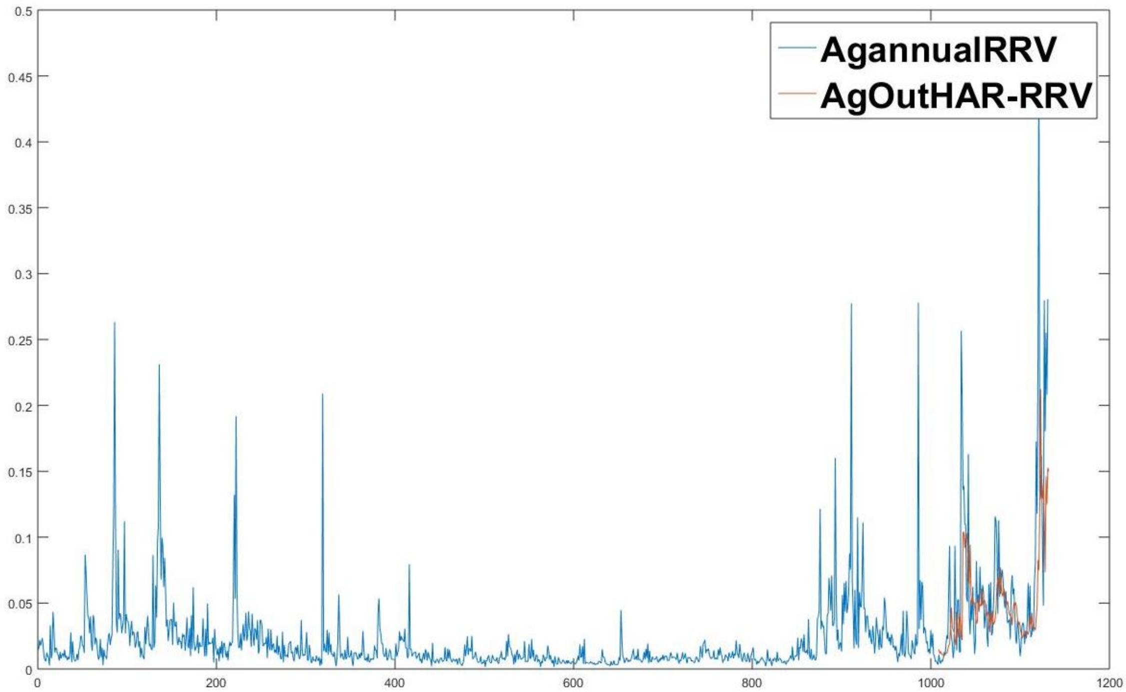

Figure 10. Model HAR-RRV Forecast for Ag(SHFE). The horizontal line is the timeline, and the perpendicular line is the volatility estimator. The forecasted period is the last 124 days. The actual values are represented by the blue line, while the forecasted values are represented by the orange line. The forecasted values capture the characteristics of to the actual values, but are not close to the actual values when there is a sharp jump.

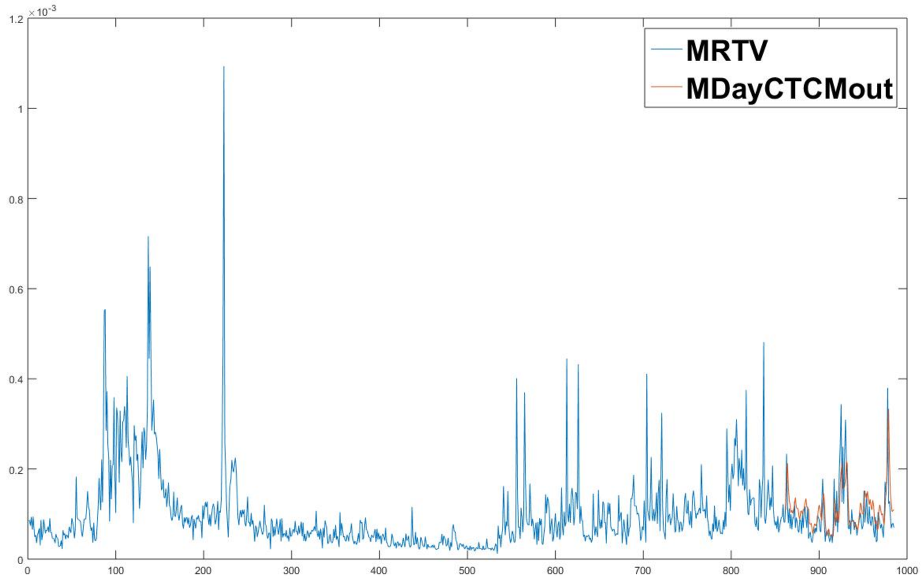

Figure 11. Model CTCM-RTV-D Forecast for M(DCE). The horizontal line is the timeline, and the perpendicular line is the volatility estimator. The forecasted period is the last 124 days. The actual values are represented by the blue line, while the forecasted values are represented by the orange line. The forecasted values are a bit higher than the actual values except for the values around the Day 920.

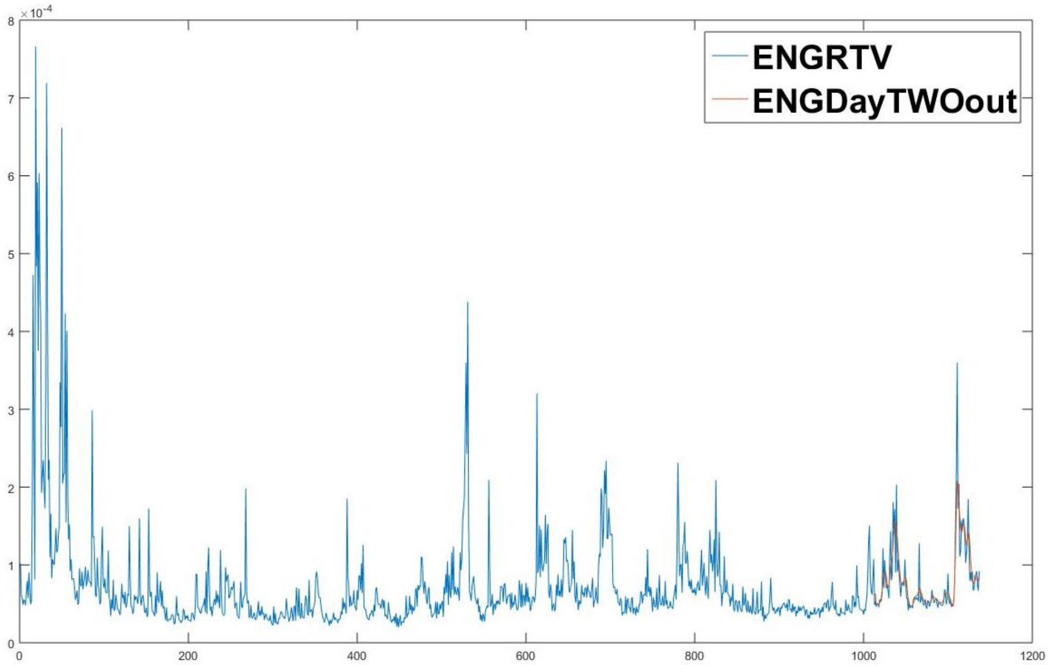

Figure 12. Model TWO-RTV-D Forecast for ENG. The horizontal line is the timeline, and the perpendicular line is the volatility estimator. The forecasted period is the last 124 days. The actual values are represented by the blue line, while the forecasted values are represented by the orange line. The forecasted values are close to the actual values, except for a sharp jump around the Day 1110.

3.3. Summary of Results

The following table (Table 8) just lists the name of the models which are most efficient.

Table 8. The most efficient models.

Products	Estimation MZ-R^2	Forecast MAPE	Forecast RMSE	Forecast Theil's U
Ag(SHFE)	CTCM-RTV-D	CTCM-RTV-D	CTCM-RTV-D	HAR-RRV
Al(SHFE)	CTCM-RTV-D	TWO-RTV-D	CTCM-RTV-D	CTCM-RTV-D
Cu(SHFE)	CTCM-RTV-D	CTCM-RTV-D	CTCM-RTV-D	HAR-RBV
I(DCE)	HAR-RTV-OI	TWO-RTV-D	CTCM-RTV-D	HAR-RTV
J(DCE)	CTCM-RTV-D	TWO-RTV-D	TWO-RTV-D	HAR-RRV-RBV
JM(DCE)	CTCM-RTV-D	TWO-RTV-D	TWO-RTV-D	HAR-RRV-OI
M(DCE)	CTCM-RTV-D	TWO-RTV-D	CTCM-RTV-D	CTCM-RTV-D
RM(CZCE)	CTCM-RTV-D	TWO-RTV-D	TWORW-RTV	CTCM-RRV-D
ENG	CTCM-RTV-D	CTCM-RTV-D	TWO-RTV-D	TWO-RTV-D
MTR	CTCM-RTV-D	CTCM-RTV-D	TWO-RTV-D	TWO-RTV-D
MDC	CTCM-RTV-D	CTCM-RTV-D	TWO-RTV-D	TWO-RTV-D
FINRE	CTCM-RTV-D	CTCM-RTV-D	TWO-RTV-D	TWO-RTV-D

4. Discussion

The empirical results show that the daily improved CTCM with RTV is the best model to describe the trend for most products, expect for I(DCE).

Based on MAPE and RMSE evaluation rules, all CTCM type models have better forecasting results than the HAR-RV type models in out of sample forecast, and RTV is the best estimator for all products in out of sample forecast.

According to Theil's U statistics, HAR-RV type models do well in Ag(SHFE), Cu(SHFE), I(DCE), J(DCE), and JM(DCE), and all of 3 volatility estimator RRV, RBV and RTV can be the most efficient estimators.

However, there is no efficient model which consistently satisfies all the 3 evaluation rules. We use modified Diebold–Mariano test to check whether the above most efficient model forecasting performances are similar or not. Experiment results show that most products' CTCM-RTV-D models have similar performance to TWO-RTV-D models, except for Al(SHFE). All the most efficient models of Ag(SHFE), Al(SHFE), Cu(SHFE), J(DCE) and JM(DCE) have different forecasting performance. All the most efficient models of RM(DCE) have the similar forecasting performance.

One reason the Theil's U statistics results show the HAR-RV type model more efficient than the CTCM type model is that the linear model will get a smaller Theil's U value easily. In this article, HAR-RV type model is linear model, while CTCM type mode is non-linear. Therefore, even if both of MAPE and RMSE of CTCM type model are smaller, HAR-RV type model may be more efficient according to Theil's U statistics. We do not suggest employing Theil's U statistics when comparing a linear model with a non-linear model.

Therefore, the improved CTCM introduced in this paper does not have better forecasting than CTCM provided by Harris [4]. Moreover, the open interest has a little influence on the performance of HAR-RV type models in Chinese futures market, so that the HAR-RV-OI type models introduced in this paper could be an alternative term to improve HAR-RV type models.

5. Conclusions

This paper introduces an improved cyclical to component model (CTCM) to describes and forecast Chinese futures markets and sector stocks. The main idea is that the volatility is impacted by a long trend component and a short-run mean-reverting component. The experiment shows that the improved CTCM is efficient enough to estimate the trend of products in Chinese futures markets and sector stocks, and is also more efficient than HAR-RV type models.

Moreover, this paper show that an appropriate volatility estimator improves the performance of models. The RTV employed in CTCM type models are better than RRV introduced by [4]. The empirical results also show that the open interest in Chinese futures markets has little influence on the trend of volatility.

Author Contributions: Original draft preparation, simulation and computation: J.W.; writing review and editing, project supervision and administration: C.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Research Development Fund (RDF) of Xi'an Jiaotong-Liverpool University, grant number RDF18-02-08 and Key Program Special Fund (KSF) of Xi'an Jiaotong-Liverpool University, grant number KSF-E-31.

Acknowledgments: We would like to thank the editor and two anonymous referees for their valuable comments as well as the suggestions of Xiaoquan Liu to improve the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CTCM	cyclical two-component model
CZCE	Zhengzhou Commodity Exchange
DCE	Dalian Commodity Exchange
HAR-RV type	Heterogeneous Autoregressive model of Realized Volatility type
MAPE	Mean Absolute Percent Error
$MZ\text{-}R^2$	the value of goodness-of-fit of Mincer–Zarnowitz Regression
OI	open interest
RBV	realized range-based bipower volatility
RMSE	root mean square error
RMV	realized range-based multipower variance
RRV	realized range (volatility)
RTV	realized range-based tripower volatility
SHFE	Shanghai Futures Exchange

References

1. Dana A.N. Modelling and estimation of volatility using ARCH/GARCH models in Jordans stock market, Jordan. *Asian J. Financ. Account.* **2016**, *8*, 152–167.
2. Brandt, M.; Jones, C. Volatility forecasting with range-based EGARCH models. *J. Bus. Econ. Stat.* **2006**, *79*, 61–74. [CrossRef]
3. Granger, C.W.; Joyeux, R. An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.* **1980**, *1*, 15–29. [CrossRef]
4. Harris, R.D.; Stoja, E.; Yilmaz, F. A cyclical model of exchange rate volatility. *J. Bank. Financ.* **2011**, *35*, 3055–3064. [CrossRef]
5. Corsi, F. A simple approximate long-memory model of realized volatility. *J. Financ. Econ.* **2009**, *7*, 174–196. [CrossRef]
6. Yang, X.; Zhou, X. *A New Liquidity Measure and Liquidity Premium*; Working Paper; Nottingham University Business School: Ningbo, China, 2018.
7. Christensen, K.; Podolskij, M. Asymptotic theory of range-based multipower variation. *J. Financ. Econ.* **2012**, *10*, 417–456. [CrossRef]
8. Parkinson, M. The extreme value method for estimating the variance of the rate of return. *J. Bus.* **1980**, *53*, 61–65. [CrossRef]
9. Alizadeh, S.; Brandt, M.W.; Diebold, F.X. Range-based estimation of stochastic volatility models. *J. Financ.* **2002**, *47*, 1047–1092. [CrossRef]
10. Andersen, T.G.; Bollerslev, T. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *Int. Econ. Rev.* **1998**, *39*, 885–905. [CrossRef]
11. Muller, U.; Dacorogna, M.; Dav, R.; Pictet, O.; Olsen, R.; Ward, J. Fractals and Intrinsic Time—A Challenge to Econometricians. In Proceedings of the 39th International AEA Conference on Real Time Econometrics, Ascona, Switzerland, 22–26 November 1993.

12. Andersen, T.G.; Bollerslev, T. Heterogeneous information arrivals and return volatility dynamics: Uncovering the long run in high frequency data. *J. Financ.* **1997**, *52*, 975–1005. [CrossRef]
13. Engle, R.; Lee, G. A permanent and transitory model of stock return volatility. In *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger.*; Engle, R., White, H., Eds.; Oxford University Press: New York, NY, USA, 1999.
14. Fei, T.; Liu, X.; Wen, C. F*orecast Stock Return Volatility with Duration-Based Volatility Estimator: Empirical Evidence from Chinese Stock Market*; Working Paper; University of Nottingham: Ningbo, China, 2018.
15. Ripple, R.D.; Moosa, I.A. The effect od maturity, trading volume and open interest on crude oil futures price range-based volatility. *Glob. Financ. J.* **2009**, *20*, 209–219. [CrossRef]
16. Hodrick, R.; Prescott, E. Post-war US business cycles: An empirical investigation. *J. Money Credit Bank.* **1997**, *29*, 1–16. [CrossRef]