

Research on the Daily Volatility Measure Considering the Impact of Overnight Variance and Time Segment in Chinese Stock Market

Yu Shi, Handong Li

School of Systems Science, Beijing Normal University, Beijing, China

Email: shiyu.yuer@foxmail.com, lhd@bnu.edu.cn

How to cite this paper: Shi, Y. and Li, H.D. (2018) Research on the Daily Volatility Measure Considering the Impact of Overnight Variance and Time Segment in Chinese Stock Market. *Journal of Mathematical Finance*, 8, 549-561.

<https://doi.org/10.4236/jmf.2018.83035>

Received: July 10, 2018

Accepted: August 4, 2018

Published: August 7, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Considering the overnight effect on the stock market, we construct a daily volatility measure that is formed by a linear combination of the three components, namely overnight volatility, morning realized volatility and afternoon realized volatility, and obtain the optimal solution in theory. An empirical work is performed for studying the daily volatility structure of Shanghai stock index and Shenzhen stock index in China's stock market by using our daily volatility measure. The empirical results show that, the daily volatility measure considering the impact of overnight variance and time segment performs better than original volatility measure.

Keywords

Overnight Volatility, Chinese Stock Market, Realized Volatility

1. Introduction

The modeling and forecasting of volatility is the basis of financial asset portfolio allocation, capital asset pricing and risk management, and has always been a hot topic in the financial field. With the improvement of high frequency data accessibility and the deepening of the research on high frequency data field, the use of high frequency data to estimate and model volatility has become a new trend in financial research [1]. Volatility measures based on nonparametric methods and using high frequency data are called Realized Volatility (RV) [2] [3]. RV is a well-known quantity that is constructed from high-frequency intra-day returns. Andersen and Bollerslev [4], Andersen *et al.* [5], Barndorff-Nielsen and Shephard [6] pointed out that the realized volatility can be estimated by continuous sampling of intra-day returns, and theoretically proved that it is a consistent es-

timate of the integrated volatility in an asset pricing process.

With the deepening of research, many scholars have found that the information in the non-trading period of stock markets which is the overnight period has a very important influence on the volatility (Hansen and Lunde [7] [8]; Taylor [9]; Ahoniemi and Lanne [10]; Oldfield [11]; Tsiakas [12]). The existing literature on stock market realized volatility has adopted several approaches to dealing with the overnight volatility. Hansen and Lunde [7] [8] firstly proposed the method of combining realized variance by optimizing the squared return weights of trading time and overnight time. Then, Ahoniemi and Lanne [10] verified the effectiveness of this method through empirical evidence. Christoffersen [13] proposed the method of adding the squared return of overnight to the daily volatility measure of trading time. Maderitsch [14] considered the impact of structural changes of asset prices and improved the volatility measure of Hansen and Lunde [7] [8]. In the study of the overnight effect of Chinese stock market on volatility, Sun [1] decomposed the realized volatility of trading time into two components: continuous path variation and jump, and combined with overnight volatility to form a daily volatility measure through the HAR-CJN model. Ma *et al.* [15] added overnight rate of returns as an explanatory variable to the high-frequency volatility model to study the impact on model prediction accuracy.

In general, the existing methods lack not only the analysis of the impact of asset volatility at different periods of the day, but also the research on the correlation and combination of different intraday volatility. As pointed out by Ahoniemi and Lanne [10], how to optimize overnight information in the context of realized volatility remains an issue that needs further study. As an emerging market, the Chinese stock market is in a stage of rapid development and continuous improvement, which presents a unique pattern of changes. Especially from 2014 to 2016, the huge volatility of the Chinese stock market has highlighted the changeability and vulnerability of the stock market. The study on the integration of overnight variance and its impact on stock price volatility can enable us to measure the volatility more accurately, deepen our understanding of stock market volatility and improve the level of risk management of Chinese stock market, thus promoting the stable and healthy development of it.

2. Realized Volatility and Overnight Effect

2.1. Realized Volatility

Let $\{p^*(t), t \geq 0\}$ represent the logarithmic pricing process of financial assets, and its generating mechanism can be expressed as a stochastic differential equation:

$$dp^*(t) = \mu(t)dt + \sigma(t)dw(t) \quad (1)$$

where $\mu(t)$ denotes the drift rate, $\sigma(t)$ denotes the volatility, and $w(t)$ denotes the standard Brownian motion.

The true volatility of p^* on the t -th day can be defined as $IV_t \equiv \int_{t-1}^t \sigma^2(t) dt$. As the integral of volatility, IV_t is also called the Integrated Volatility of the t -th day, and $r_t \equiv p^*(t) - p^*(t-1)$ is defined as the intra-day return of the t -th day.

Andersen and Bollerslev proposed the definition of Realized Volatility (RV). Let $M + 1$ denotes the number of price observation values at equal intervals of the day, that is, $p_i(t-1), p_i\left(t-1 + \frac{1}{M}\right), \dots, p_i(t)$. Then, there are a total of M returns in a day, and the intra-day return of the j th observation period is defined as $r_{i,t,j} \equiv p_i\left(t-1 + \frac{j}{M}\right) - p_i\left(t-1 + \frac{j-1}{M}\right), j = 1, 2, \dots, M$. The realized volatility is defined as:

$$RV_{i,t} \equiv \sum_{j=1}^M r_{i,t,j}^2. \quad (2)$$

It has been proved by Barndorff and Shephard that, without considering the jumps of asset prices, according to the Quadratic Variation theory, when $M \rightarrow \infty$, $RV_{i,t}$ converges to the integrated volatility in probability, that is, $\lim_{M \rightarrow \infty} RV_{i,t} = \int_{t-1}^t \sigma^2(t) dt$. In other words, if the sampling frequency of intra-day returns is high enough, $RV_{i,t}$ can be regarded as the consistent estimator of real volatility.

2.2. The Overnight Effect of China's Stock Market Volatility

Transactions in the Chinese stock market are concentrated on the Shanghai Stock Exchange (SSE) and the Shenzhen Stock Exchange (SZSE). The opening hours of each day are from 9:30 to 11:30 and from 13:00 to 15:00, namely, there are 4 hours of trading on each day. However, since the asset prices are changing all the time, using the price changing information observed in only four hours to describe the price changes of the whole day is inaccurate. So it is necessary to consider the price changes at the non-trading hours.

We define the intra-day return as the difference between the logarithm of the daily closing price and the logarithm of the previous daily closing price. The time of a day corresponds to the closing time of the previous day to the closing time of the day. Hence, we can divide the intra-day return into four periods:

Phase I. Overnight period from 15:00 on the previous day to 9:30 on the day. The overnight return of a stock is defined as the difference between the logarithm of the day's opening price at 9:30 and the logarithm of the previous day's closing price at 15:00, represented by r_{1t} ;

Phase II. Opening hours from 9:30 to 11:30. The morning return of a stock can be obtained in the same way as calculating the intra-day return, represented by r_{2t} ;

Phase III. Lunch break from 11:30 to 13:00. The stock's midday return is defined as the difference between the logarithm of the day's opening price at 13:00 and the logarithm of the day's closing price at 11:30, represented by r_{3t} ;

Phase IV. Opening hours from 13:00 to 15:00. The afternoon return of a stock

can be obtained in the same way as calculating the intra-day return, represented by r_{4t} .

The volatilities of overnight and midday returns can be expressed as the square of the returns, that is, r_{1t}^2 and r_{3t}^2 . The volatilities of morning and afternoon returns are represented by realized volatility, that is, RV_{2t} and RV_{4t} . The partition results of the four time periods are shown in **Figure 1**.

To compare the changes of daily volatility at different periods, we use the 1-minute high-frequency data of Shanghai Composite Index and Shenzhen Component Index from January 2, 2014 to November 2, 2015 for comparative analysis, including 429 days of valid data of Shanghai Composite Index and 437 days of valid data of Shenzhen Component Index. Due to the partial missing and abnormal values of 1-minute data of Shanghai Composite Index obtained, in order to ensure that the analysis results are in line with the actual situation, we firstly preprocess the data by interpolating. To eliminate the influence of micro-structure noise of high-frequency data, we use high-frequency data at intervals of 5, 10, 15, 20 and 30 minutes respectively for analysis. The results show that the statistical characteristics of the volatilities of the returns at different time intervals display roughly the same variation pattern. **Table 1** and **Table 2** respectively show the statistical characteristics of realized volatility obtained at a 5-minute sampling interval.

It can be seen from **Table 1** and **Table 2** that both Shanghai Composite Index and Shenzhen Composite Index have large volatilities of the morning returns. Volatilities of the afternoon returns are smaller than that of the morning,

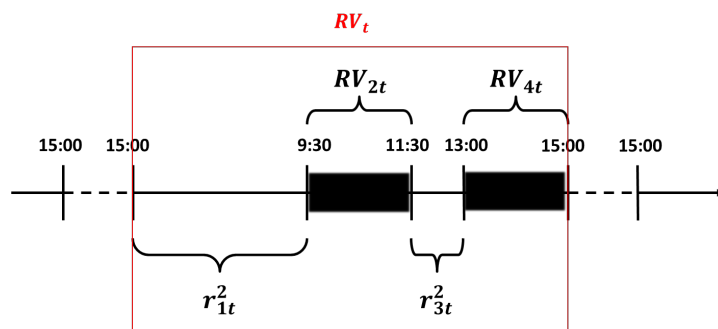


Figure 1. The decomposition of the daily volatility structure of Chinese stock market.

Table 1. Daily volatility statistical characteristics of Shanghai Composite Index at 5-minute sampling interval.

Statistical Characteristics	Mean	Variance
r_{1t}^2	1.30e-04	1.43e-07
RV_{2t}	1.85e-04	1.02e-07
r_{3t}^2	4.82e-06	5.95e-10
RV_{4t}	1.46e-04	7.21e-08

Source: Data from Shanghai Stock Exchange.

Table 2. Daily volatility statistical characteristics of Shenzhen Component Index at 5-minute sampling interval.

Statistical Characteristics	Mean	Variance
r_{1t}^2	1.13e-04	2.41e-07
RV_{2t}	1.70e-04	1.02e-07
r_{3t}^2	1.93e-07	3.00e-12
RV_{4t}	1.45e-04	7.71e-08

Source: Data from Shenzhen Stock Exchange.

which is basically consistent with the reality of stock markets. The volatilities of the overnight period are slightly less than that of the afternoon, but are in the same order of magnitude as the averaged volatilities over the trading period. Hence, it can be seen that the overnight effect is obvious and should not be ignored. Because the volatility of midday returns is extremely small compared to that of the other three time periods, in the following analysis we choose overnight volatility, morning realized volatility and afternoon realized volatility as the main components of daily volatility.

3. Research Methods

The intra-day return of the stock market is defined as the difference between the logarithm of the day's closing price and the logarithm of the previous day's closing price. According to the division of the four time periods in **Figure 1**, the intra-day return of the t -th trading day of the stock market can be defined as $r_t \equiv r_{1t} + r_{2t} + r_{3t} + r_{4t}$, where r_{2t}, r_{4t} are the morning and afternoon returns of the opening hours and the volatility can be expressed by the realized volatility, which is the sum of the squares of returns on trading time, that is, $RV_{2t} = \sum_{i=1}^m r_{2t-i}^2, RV_{4t} = \sum_{i=1}^m r_{4t-i}^2$. And $r_{2t-i}, r_{4t-i}, i = 1, \dots, m$ respectively represent the morning and afternoon returns of m equal time intervals of the opening hours. Compared with the four periods of intra-day returns, the implied integrated volatility can be divided into four parts, namely,

$$IV_t = IV_{1t} + IV_{2t} + IV_{3t} + IV_{4t}.$$

To simplify writing, we define $E_t(\cdot) = E[\cdot | IV_t]$ to represent conditional expectations. The basic hypotheses discussed below are:

Assumption I. $E_t[IV_{it}] = \delta_i IV_t, i = 1, 2, 3, 4;$

Assumption II. $E_t[\omega_i RV_{it} - IV_{it}] = 0, i = 2, 4;$

Assumption III. $E_t[\omega_i r_{it}^2 - IV_{it}] = 0, i = 1, 3;$

Assumption IV. $Cov(IV_{it}, IV_{jt}) = 0, i \neq j, i, j = 1, 2, 3, 4;$

where δ_i, ω_i are constants, respectively.

Assumption I shows that the expected value of the integrated volatility of each time period is fixed in proportion to integrated volatility of the whole day. Assumption II requires that the conditional deviation rates of the volatilities of the morning and afternoon returns are proportional to the integrated volatilities of

the corresponding period. Assumption III requires that overnight and midday volatilities are also proportional to the integrated volatilities of the corresponding period. Assumption IV means that the integrated volatilities at different periods are irrelevant. Assumption IV is supposed to simplify the problem, which is generally not true in practical problems.

To give a more accurate description of the volatility in stock market, based on the model of Hansen and Lunde, we subdivide the opening hours into morning and afternoon and add in volatility of overnight returns. Hence, we define the daily volatility measure of the Chinese stock market as a linear combination of overnight volatility, morning realized volatility and afternoon realized volatility:

$$RV_t \equiv \theta_1 r_t^2 + \theta_2 RV_{2t} + \theta_4 RV_{4t} \tag{3}$$

where $\theta \equiv (\theta_1, \theta_2, \theta_4)'$ is the parameter to be estimated.

To facilitate this discussion, we define unconditional expectations $\mu \equiv E[IV_t]$, $\mu_1 \equiv E[r_t^2]$, $\mu_2 \equiv E[RV_{2t}]$, $\mu_4 \equiv E[RV_{4t}]$. Then we have the following conclusions.

Theorem 1. For all $\theta \equiv (\theta_1, \theta_2, \theta_4)'$ that satisfy $\theta_1\mu_1 + \theta_2\mu_2 + \theta_4\mu_4 = \mu$, there is $E_t[RV_t] = IV_t$.

Proof of Theorem 1: see Appendix.

The purpose of the established volatility measure RV_t is to better describe integrated volatility IV_t , and the difference between them can be measured by the mean square error. So this problem turns into the following optimization problem.

$$\min_{\theta \in \Theta} \{E[RV_t - IV_t]^2\} \tag{4}$$

where Θ is the value interval of parameter $\theta = (\theta_1, \theta_2, \theta_4)'$.

But in fact, since IV_t is the integral of instantaneous volatility, it is an unobservable potential variable. Moreover, the asset price itself is disturbed by the micro-structure noise, so we cannot get a more accurate measurement value of actual IV_t , and other methods need to be considered. Then we give the following theorem.

Theorem 2. For $\theta \in \Theta$, if $E_t[RV_t - IV_t] = 0$, then $\min_{\theta \in \Theta} \{E[RV_t - IV_t]^2\}$ is equivalent to $\min_{\theta \in \Theta} D[RV_t]$, where $D[RV_t]$ is the variance of RV_t .

Proof of Theorem 2: see Appendix.

From Theorem 2, the original problem can be transformed into an optimization problem:

$$\min_{\theta \in \Theta} D[RV_t] = \min_{\theta \in \Theta} D(\theta_1 r_t^2 + \theta_2 RV_{2t} + \theta_4 RV_{4t}) \text{ s.t. } \theta_1\mu_1 + \theta_2\mu_2 + \theta_4\mu_4 = \mu$$

Hence, we have the following conclusions.

Theorem 3. The solution of the optimization problem $\min_{\theta \in \Theta} D(\theta_1 r_t^2 + \theta_2 RV_{2t} + \theta_4 RV_{4t}) \text{ s.t. } \theta_1\mu_1 + \theta_2\mu_2 + \theta_4\mu_4 = \mu$ is

$$\hat{\theta}_1 = \alpha \frac{\mu}{\mu_1}; \hat{\theta}_2 = \beta \frac{\mu}{\mu_2}; \hat{\theta}_4 = (1 - \alpha - \beta) \frac{\mu}{\mu_4} \tag{5}$$

where

$$\alpha = \frac{\frac{\sigma_{33} - \sigma_{13}}{\mu_4^2} - \frac{\sigma_{13}}{\mu_1\mu_4}}{\frac{\sigma_{11} + \sigma_{33}}{\mu_1^2} + \frac{\sigma_{33}}{\mu_4^2} - 2\frac{\sigma_{13}}{\mu_1\mu_4}} + \left[\frac{\frac{\sigma_{13} + \sigma_{23}}{\mu_1\mu_4} - \frac{\sigma_{12}}{\mu_2\mu_4} - \frac{\sigma_{33}}{\mu_4^2}}{\frac{\sigma_{11} + \sigma_{33}}{\mu_1^2} + \frac{\sigma_{33}}{\mu_4^2} - 2\frac{\sigma_{13}}{\mu_1\mu_4}} \right]$$

$$\times \left[\frac{\left(\frac{\sigma_{23} - \sigma_{13}}{\mu_2\mu_4} - \frac{\sigma_{13}}{\mu_1\mu_4} \right) \left(\frac{\sigma_{11} + \sigma_{33}}{\mu_1^2} - 2\frac{\sigma_{13}}{\mu_1\mu_4} \right) + \left(\frac{\sigma_{11} - \sigma_{13}}{\mu_1^2} - \frac{\sigma_{12}}{\mu_1\mu_4} + \frac{\sigma_{23}}{\mu_2\mu_4} \right) \left(\frac{\sigma_{13} - \sigma_{33}}{\mu_1\mu_4} - \frac{\sigma_{33}}{\mu_4^2} \right)}{\left(\frac{\sigma_{13} + \sigma_{23}}{\mu_1\mu_4} - \frac{\sigma_{12}}{\mu_2\mu_4} - \frac{\sigma_{33}}{\mu_4^2} \right) \left(\frac{\sigma_{11} - \sigma_{13}}{\mu_1^2} - \frac{\sigma_{12}}{\mu_1\mu_4} + \frac{\sigma_{23}}{\mu_2\mu_4} \right) - \left(\frac{\sigma_{11} + \sigma_{33}}{\mu_1^2} - 2\frac{\sigma_{13}}{\mu_1\mu_4} \right) \left(\frac{\sigma_{22} - \sigma_{12}}{\mu_2^2} - \frac{\sigma_{23}}{\mu_2\mu_4} + \frac{\sigma_{13}}{\mu_1\mu_4} \right)} \right]$$

$$\beta = \frac{\left(\frac{\sigma_{23} - \sigma_{13}}{\mu_2\mu_4} - \frac{\sigma_{13}}{\mu_1\mu_4} \right) \left(\frac{\sigma_{11} + \sigma_{33}}{\mu_1^2} - 2\frac{\sigma_{13}}{\mu_1\mu_4} \right) + \left(\frac{\sigma_{11} - \sigma_{13}}{\mu_1^2} - \frac{\sigma_{12}}{\mu_1\mu_4} + \frac{\sigma_{23}}{\mu_2\mu_4} \right) \left(\frac{\sigma_{13} - \sigma_{33}}{\mu_1\mu_4} - \frac{\sigma_{33}}{\mu_4^2} \right)}{\left(\frac{\sigma_{13} + \sigma_{23}}{\mu_1\mu_4} - \frac{\sigma_{12}}{\mu_2\mu_4} - \frac{\sigma_{33}}{\mu_4^2} \right) \left(\frac{\sigma_{11} - \sigma_{13}}{\mu_1^2} - \frac{\sigma_{12}}{\mu_1\mu_4} + \frac{\sigma_{23}}{\mu_2\mu_4} \right) - \left(\frac{\sigma_{11} + \sigma_{33}}{\mu_1^2} - 2\frac{\sigma_{13}}{\mu_1\mu_4} \right) \left(\frac{\sigma_{22} - \sigma_{12}}{\mu_2^2} - \frac{\sigma_{23}}{\mu_2\mu_4} + \frac{\sigma_{13}}{\mu_1\mu_4} \right)}$$

and $\sigma_{11} \equiv D(r_{1t}^2), \sigma_{22} \equiv D(RV_{2t}), \sigma_{33} \equiv D(RV_{4t}), \sigma_{12} \equiv Cov(r_{1t}^2, RV_{2t}),$
 $\sigma_{13} \equiv Cov(r_{1t}^2, RV_{4t}), \sigma_{23} \equiv Cov(RV_{2t}, RV_{4t}).$

In particular, when Assumption IV is true, that is, $\sigma_{12} \equiv Cov(r_{1t}^2, RV_{2t}) = 0,$
 $\sigma_{13} \equiv Cov(r_{1t}^2, RV_{4t}) = 0, \sigma_{23} \equiv Cov(RV_{2t}, RV_{4t}) = 0,$ then we have

$$\hat{\theta}_1 = \frac{\mu\mu_1\sigma_{22}\sigma_{33}}{\mu_4^2\sigma_{11}\sigma_{22} + \mu_1^2\sigma_{22}\sigma_{33} + \mu_2^2\sigma_{11}\sigma_{33}}$$

$$\hat{\theta}_2 = \frac{\mu\mu_2\sigma_{11}\sigma_{33}}{\mu_4^2\sigma_{11}\sigma_{22} + \mu_1^2\sigma_{22}\sigma_{33} + \mu_2^2\sigma_{11}\sigma_{33}} \tag{6}$$

$$\hat{\theta}_4 = \frac{\mu - \hat{\theta}_1\mu_1 - \hat{\theta}_2\mu_2}{\mu_4}$$

Proof of Theorem 3: see Appendix.

4. Empirical Analysis

4.1. Selection of Sample Data

The empirical analysis data used in this paper is still 1-minute closing price data of China’s stock market index from January 2, 2014 to November 2, 2015. Since most studies have confirmed that the 5-minute data can be considered almost impervious to the micro-structure noise of high-frequency data, and we have analyzed the 5, 10, 15, 20, 30 minutes intervals and find that there is roughly the same variation pattern, we mainly analyze Shanghai Composite Index and Shenzhen Component Index with a sampling interval of 5 minutes in the following empirical part. After processing the high-frequency data at a sampling interval of 5 minutes, 50 pieces of high-frequency data can be recorded every trading day. Then we have a total of 21,450 pieces of high-frequency data covering 429 trading days of Shanghai Composite Index, and 21,850 pieces of high-frequency data covering 437 trading days of Shenzhen Component Index.

4.2. Correlation Analysis of Returns and Volatilities

In Assumption IV, we assume that the volatilities of different time periods are

irrelevant. Since this is a very strong assumption, we need to firstly test the correlation between returns and volatilities in practical applications.

We conduct an autocorrelation analysis of 5, 10, 15, 20 and 30 minutes return series of the two stock indexes. The results show that there is no autocorrelation in the series of intraday returns within the above time intervals. Then we conduct correlation tests on the series of the overnight volatility r_1^2 , the morning realized volatility RV_2 and the afternoon realized volatility RV_4 . The results are shown in **Table 3** and **Table 4**.

As can be seen from the sample correlation coefficients in **Table 3** and **Table 4**, the overnight volatility is correlated with both the morning realized volatility and the afternoon realized volatility. In this case, the existence of correlation cannot be ignored. Hence, we use Equation (5) to calculate the daily volatility.

4.3. Parameter Estimation and Comparison of Volatility Measures

Using the result of Theorem 3, we calculate the parameters of 5, 10, 15, 20 and 30 minutes sampling intervals. As the final results obtained from different intervals do not differ much, we chose the 5-minute interval as the representative for the following analysis and explanation. Bring the return series into the formula in Theorem 3, we calculate the parameters as shown in **Table 5**.

From the formula $RV_t = \theta_1 r_{1t}^2 + \theta_2 RV_{2t} + \theta_4 RV_{4t}$, we can get the daily volatility measure that we construct (expressed as XRV for convenience of differentiation). We take the square of the intra-day returns r^2 as a reflection of the real volatility, and use RV to represent the daily volatility measure which is constructed only by using high-frequency data of opening hours and without taking the overnight effect into account. We calculate the mean and variance of these

Table 3. Sample correlation coefficients for overnight volatility, morning realized volatility and afternoon realized volatility of Shanghai Composite Index.

Sample Correlation Coefficient	r_1^2	RV_2	RV_4
r_1^2	1.0000		
RV_2	0.3675	1.0000	
RV_4	0.2319	0.7187	1.0000

Source: Data from Shanghai Stock Exchange.

Table 4. Sample correlation coefficients for overnight volatility, morning realized volatility and afternoon realized volatility of Shenzhen Component Index.

Sample Correlation Coefficient	r_1^2	RV_2	RV_4
r_1^2	1.0000		
RV_2	0.4034	1.0000	
RV_4	0.2840	0.6942	1.0000

Source: Data from Shenzhen Stock Exchange.

Table 5. Results of parameters at a 5-minute sampling interval.

Parameter	Shanghai Composite Index	Shenzhen Component Index
μ_1	1.13e-04	1.30e-04
μ_2	1.70e-04	1.85e-04
μ_4	1.45e-04	1.46e-04
σ_{11}	2.41e-07	1.43e-07
σ_{22}	1.02e-07	1.02e-07
σ_{44}	7.71e-08	7.21e-08
σ_{12}	0.3675	0.4034
σ_{14}	0.2319	0.2840
σ_{24}	0.7187	0.6942
$\hat{\theta}_1$	0.4163	0.5803
$\hat{\theta}_2$	1.1420	1.0562
$\hat{\theta}_4$	1.2902	1.3006

three volatility measures of Shanghai Composite Index and Shenzhen Component Index respectively, and show the statistical characteristics in **Table 6**.

It can be seen that the daily volatility measure XRV constructed by us is closer to the real volatility value than the realized volatility measure RV which does not consider the overnight effect.

Finally, we use the mean square error (MSE) as the standard to measure the error, which is the most commonly used form of loss function in such judgment. According to the principle of least square method, the smaller the sum of squared residuals, the more consistent the estimated value is with the real value. In practical application, MSE is usually used as a measurement index. The smaller the expected value of the squared difference between the estimated value and the real value, the more accurate the model is. The specific definition is $MSE = \frac{1}{n} \sum_{i=1}^n (\tilde{Y}_i - Y_i)^2$, and MSEs between the volatility measures and the real volatility are shown in **Table 7**.

As we can see from **Table 7**, the mean square errors of the daily volatility measure we build (XRV) are less than the corresponding values of the measure without considering the overnight variance and time segment (RV).

4.4. Summary

Based on the above analysis results, we can draw the conclusion that, the daily volatility measure considering the impact of overnight variance and time segment (XRV) is superior to the realized volatility (RV), and can reflect the real situation of volatility more comprehensively.

Table 6. Statistical characteristics of volatility measures at a 5-minute sampling interval.

Stock Index	Statistical characteristics	Mean	Variance
Shanghai Composite Index	XRV	4.29e-04	5.81e-07
	RV	3.15e-04	3.07e-07
	r^2	3.90e-04	9.17e-07
Shenzhen Component Index	XRV	4.62e-04	5.53e-07
	RV	3.36e-04	3.09e-07
	r^2	4.57e-04	9.32e-06

Table 7. The mean square error between the volatility measures and the real volatility at a 5-minute sampling interval.

Stock Index	Volatility Measure	MSE
Shanghai Composite Index	XRV	7.99e-07
	RV	8.16e-07
Shenzhen Component Index	XRV	7.14e-07
	RV	7.77e-07

5. Conclusions

Due to the use of more intra-day data, the realized volatility measure based on high-frequency return series shows better statistical properties than parametric model in characterizing historical volatility. Since the trading hours of stock markets only account for a small part of a day, which are divided into two periods of morning and afternoon, asset prices change continuously. Hence, the realized volatility composed of the return series of trading time cannot fully characterize the daily volatility. Base on this point, the main work of this paper is to establish the optimized realized volatility statistics through the analyzing and processing of high-frequency trading data of China's stock market, and compare it with the original measure through mean square error to judge the pros and cons of the new volatility measure. The main empirical results show that in terms of Shanghai Composite Index and Shenzhen Component Index, the daily volatility measure considering the impact of overnight variance and time segment is superior to realized volatility measure without considering them.

This paper proposes a daily volatility measure for Chinese stock market, which considers the impact of overnight variance and time segment. This approach is helpful for us to better understand the volatility structure of Chinese stock market and give a more accurate measure of volatility. Since the high-frequency return series are affected by the microstructure noise, and there are jumps in asset prices when it changes continuously, these factors will make the realized volatility measure have certain deviations. Therefore, the improvement of volatility measure based on microstructure and price jumps is a further work direction in the future.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Sun, J. (2014) Modeling and Forecasting the Volatility of China Stock Market Considering the Impact of Jump and Overnight Variance. *Chinese Journal of Management Science*, **22**, 114-124.
- [2] McAleer, M. and Medeiros, M.C. (2008) Realized Volatility: A Review. *Econometric Reviews*, **27**, 10-45. <https://doi.org/10.1080/07474930701853509>
- [3] Andersen, T.G. (2003) Modeling and Forecasting Realized Volatility. *Econometrica*, **71**, 579-625. <https://doi.org/10.1111/1468-0262.00418>
- [4] Andersen, T.G. and Bollerslev, T. (1998) Deutsche Mark-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies. *Journal of Finance*, **53**, 219-265. <https://doi.org/10.1111/0022-1082.85732>
- [5] Andersen, T.G., Bollerslev, T. and Huang, X. (2011) A Reduced Form Framework for Modeling Volatility of Speculative Prices Based on Realized Variation Measures ☆. *Journal of Econometrics*, **160**, 176-189. <https://doi.org/10.1016/j.jeconom.2010.03.029>
- [6] Barndorff-Nielsen, O.E. and Shephard, N. (2002) Estimating Quadratic Variation Using Realized Variance. *Journal of Applied Econometrics*, **17**, 457-477. <https://doi.org/10.1002/jae.691>
- [7] Hansen, P.R. and Lunde, A. (2005) A Realized Variance for the Whole Day Based on Intermittent High-Frequency Data. *Journal of Financial Econometrics*, **3**, 525-554. <https://doi.org/10.1093/jjfinec/nbi028>
- [8] Hansen, P.R. and Lunde, A. (2003) An Optimal and Unbiased Measure of Realized Variance Based on Intermittent High-Frequency Data.
- [9] Taylor, N. (2007) A Note on the Importance of Overnight Information in Risk Management Models. *Journal of Banking & Finance*, **31**, 161-180. <https://doi.org/10.1016/j.jbankfin.2006.01.004>
- [10] Ahoniemi, K. and Lanne, M. (2013) Overnight Stock Returns and Realized Volatility. *International Journal of Forecasting*, **29**, 592-604. <https://doi.org/10.1016/j.ijforecast.2013.03.006>
- [11] Oldfield, G.S. and Rogalski, R.J. (1980) A Theory of Common Stock Returns over Trading and Non-Trading Periods. *Journal of Finance*, **35**, 729-751. <https://doi.org/10.1111/j.1540-6261.1980.tb03495.x>
- [12] Tsiakas, I. (2008) Overnight Information and Stochastic Volatility: A Study of European and US Stock Exchanges. *Journal of Banking & Finance*, **32**, 251-268. <https://doi.org/10.1016/j.jbankfin.2007.03.008>
- [13] Christoffersen, P., Feunou, B., Jacobs, K. and Meddahi, N. (2012) The Economic Value of Realized Volatility. Peter F Christoffersen.
- [14] Maderitsch, R. (2017) 24-Hour Realized Volatilities and Transatlantic Volatility Interdependence. *Quantitative Finance*, **17**, 1-13. <https://doi.org/10.1080/14697688.2016.1206958>
- [15] Ma, F., Wei, Y., Huang, S.D. and Zhuang, X.Y. (2016) Can Overnight Returns Improve the Forecasting Performance of High-Frequency Volatility Models. *Journal of Systems Engineering*, **31**, 783-797.

Appendix

1) Proof of Theorem 1:

By hypothesis, there is

$$\begin{aligned} E_t [RV_t] &= E_t \left[\theta_1 r_{it}^2 + \theta_2 RV_{2t} + \theta_4 RV_{4t} \right] \\ &= \theta_1 \left(\frac{1}{\omega_1 \delta_1} \right) IV_t + \theta_2 \left(\frac{1}{\omega_2 \delta_2} \right) IV_t + \theta_4 \left(\frac{1}{\omega_4 \delta_4} \right) IV_t \\ &= \left[\theta_1 \left(\frac{1}{\omega_1 \delta_1} \right) + \theta_2 \left(\frac{1}{\omega_2 \delta_2} \right) + \theta_4 \left(\frac{1}{\omega_4 \delta_4} \right) \right] IV_t \end{aligned}$$

and by the law of total expectation, we have

$$E [RV_t] = E \left[E_t (RV_t) \right] = \left[\theta_1 \left(\frac{1}{\omega_1 \delta_1} \right) + \theta_2 \left(\frac{1}{\omega_2 \delta_2} \right) + \theta_4 \left(\frac{1}{\omega_4 \delta_4} \right) \right] \mu$$

and

$$E [RV_t] = E \left[\theta_1 r_{it}^2 + \theta_2 RV_{2t} + \theta_4 RV_{4t} \right] = \theta_1 \mu_1 + \theta_2 \mu_2 + \theta_4 \mu_4 = \mu$$

then we have

$$\theta_1 \left(\frac{1}{\omega_1 \delta_1} \right) + \theta_2 \left(\frac{1}{\omega_2 \delta_2} \right) + \theta_4 \left(\frac{1}{\omega_4 \delta_4} \right) = 1$$

so $E_t [RV_t] = IV_t$.

2) Proof of Theorem 1:

We define $D_t(\cdot) = D[\cdot | IV_t]$, and by conditional expectation formula,

$$E_t [RV_t - IV_t]^2 = D_t [RV_t - IV_t] + [E_t (RV_t - IV_t)]^2 = D_t [RV_t]$$

Therefore, by the law of total expectation, we can have

$$E (RV_t - IV_t)^2 = E \left\{ E_t [RV_t - IV_t]^2 \right\} = E \left\{ D_t [RV_t] \right\} = D [RV_t]$$

3) Proof of Theorem 3:

Consider the minimum variance of the linear combination

$\alpha X_1 + \beta X_2 + (1 - \alpha - \beta) X_3$ of random variables X_1, X_2, X_3 , that is,

$$\min_{\alpha, \beta} D \left[\alpha X_1 + \beta X_2 + (1 - \alpha - \beta) X_3 \right].$$

Let

$$\begin{aligned} F(\alpha, \beta) &= D \left[\alpha X_1 + \beta X_2 + (1 - \alpha - \beta) X_3 \right] \\ &= \alpha^2 DX_1 + \beta^2 DX_2 + (1 - \alpha - \beta) DX_3 \\ &\quad + 2\alpha\beta \text{Cov}(X_1, X_2) + 2\alpha(1 - \alpha - \beta) \text{Cov}(X_1, X_3) \\ &\quad + 2\beta(1 - \alpha - \beta) \text{Cov}(X_2, X_3) \end{aligned}$$

To simplify the calculation, let $c_{11} \equiv DX_1, c_{22} \equiv DX_2, c_{33} \equiv DX_3,$
 $c_{12} \equiv \text{Cov}(X_1, X_2), c_{13} \equiv \text{Cov}(X_1, X_3), c_{23} \equiv \text{Cov}(X_2, X_3)$. Take the partial derivative with respect to α and β respectively, and set the partial derivative to 0, then we have

$$\alpha = \frac{c_{33} - c_{13}}{c_{11} + c_{33} - 2c_{13}} + \left[\frac{c_{13} + c_{23} - c_{12} - c_{33}}{c_{11} + c_{33} - 2c_{13}} \right]$$

$$\times \left[\frac{(c_{23} - c_{13})(c_{11} + c_{33} - 2c_{13}) + (c_{11} - c_{13} - c_{12} + c_{23})(c_{13} - c_{33})}{(c_{13} + c_{23} - c_{12} - c_{33})(c_{11} - c_{13} - c_{12} + c_{23}) - (c_{11} + c_{33} - 2c_{13})(c_{22} - c_{12} - c_{23} + c_{13})} \right]$$

$$\beta = \frac{(c_{23} - c_{13})(c_{11} + c_{33} - 2c_{13}) + (c_{11} - c_{13} - c_{12} + c_{23})(c_{13} - c_{33})}{(c_{13} + c_{23} - c_{12} - c_{33})(c_{11} - c_{13} - c_{12} + c_{23}) - (c_{11} + c_{33} - 2c_{13})(c_{22} - c_{12} - c_{23} + c_{13})}$$

By the condition $\theta_1\mu_1 + \theta_2\mu_2 + \theta_4\mu_4 = \mu$ and normalizing it, we have

$$\theta_1 \frac{\mu_1}{\mu} + \theta_2 \frac{\mu_2}{\mu} + \theta_4 \frac{\mu_4}{\mu} = 1. \text{ And let } X_1 = \frac{\mu}{\mu_1} r_{1t}^2, X_2 = \frac{\mu}{\mu_2} RV_{2t}, X_3 = \frac{\mu}{\mu_4} RV_{4t}, \text{ then}$$

$$\text{we have } c_{11} = \frac{\mu^2}{\mu_1^2} \sigma_{11}, c_{22} = \frac{\mu^2}{\mu_2^2} \sigma_{22}, c_{33} = \frac{\mu^2}{\mu_4^2} \sigma_{33}, c_{12} = \frac{\mu^2}{\mu_1\mu_2} \sigma_{12}, c_{13} = \frac{\mu^2}{\mu_1\mu_4} \sigma_{13},$$

$$c_{23} = \frac{\mu^2}{\mu_2\mu_4} \sigma_{23}. \text{ Bring them into the above results and we have}$$

$$\hat{\theta}_1 = \alpha \frac{\mu}{\mu_1}; \hat{\theta}_2 = \beta \frac{\mu}{\mu_2}; \hat{\theta}_4 = (1 - \alpha - \beta) \frac{\mu}{\mu_4}$$

and we can get α and β .