PressAcademia
www.pressacademia.org/journals/jefa
JEFA
ISSN 2148-6697
Journal of Economics, Finance and Accounting YEAR 2025 VOLUME 12 ISSUE 1

UTILIZING THE INFORMATION CONTENT OF TRADING AND NON-TRADING PERIODS INCLUDING LUNCH BREAKS FOR STOCK MARKET VOLATILITY FORECASTING

DOI: 10.17261/Pressacademia.2025.1962
JEFA- V.12-ISS.1-2025(1)-p.1-9

Tseng-Chan Tseng[1], Chih Huang[2]
[1]Zhaoqing University, Economics and Management College, Zhaoqing, China.
tctseng6368@gmail.com , ORCID: 0000-0001-5259-817X
[2]Shaoguan University, Business School, Shaoguan, China.
2753426367@qq.com , ORCID: 0000-0002-08227723

Date Received: February 1, 2025 **Accepted:** May 4, 2025

To cite this document
Tseng and Huang (2025). Utilizing the information content of trading and non-trading periods including lunch breaks for stock market volatility forecasting. Journal of Economics, Finance and Accounting (JEFA), 12(1), 1-9.
Permanent link to this document: http://doi.org/10.17261/Pressacademia.2025.1962

ABSTRACT

Purpose- This study investigates the empirical effects of information dissemination dynamics across active trading sessions and market closures on Chinese stock market volatility.
Methodology- This paper uses intraday data to explore the influences of information transmission during trading and non-trading periods (including lunch breaks that divide each trading day into two distinct sessions) on volatility in China's stock markets, and to forecast such volatility through modelling.
Findings- Its findings demonstrate that absolute overnight return and positive lunch break return both play important roles in future volatility. Moreover, the empirical results suggest that morning-session RRV is positively linked with volatility over longer prediction horizons, and afternoon-session RRV, with volatility over shorter ones.
Conclusion- Finally, this paper proposes that a simplified model, which only considers morning-session RRV, can improve the accuracy of prediction of future realized volatility.

Keywords: Overnight returns, trading volume, lunch break returns, morning and afternoon sessions, realized range-based volatility.
JEL Codes: C32, C51, G10

1. INTRODUCTION

The process of information transmission is the main reason for price fluctuation in stock markets. The sequential information arrival (SIA) hypothesis, originally proposed by Copeland (1976) and extended by Jennings et al. (1981), holds that the new information received by traders arrives in a sequential, random fashion, and assumes that there will be intermediate equilibria en route to a final state of equilibrium. Subsequently, Andersen (1996) both suggested that trading volume can serve as a proxy measure of the unobservable amount of information that flows into the market. Darrat et al. (2003) supported the SIA hypothesis, on the grounds that trading volume and return volatility followed a clear lead-lag pattern in a large number of the Dow Jones Industrial Average stocks.

However, information transmission in non-trading periods will also cause changes in stock prices, especially opening ones. In particular, overnight returns are often regarded as an information-transmission variable. Gallo (2001) demonstrated the explanatory power of overnight returns over the squared residuals of intra-day returns and also found that the former provided a better fit for conditional-variance estimates. Ahoniemi and Lanne (2013) found that a realized volatility estimator that optimally incorporates overnight information is more accurate in-sample for the S&P 500 index. Todorova and Souček (2014) showed that considering overnight information separately, rather than adding it to daily realized-volatility estimates, consistently led to better out-of-sample results despite the higher number of parameters involved. And some prior studies have suggested that overnight returns play an important role in the volatility of asset returns (Wang et al., 2015; Jayawardena et al., 2016; Zhu et al., 2017; Jayawardena et al., 2020; Kim and Goh, 2024; Masyhuri et al., 2024).

On the other hand, despite Ahoniemi and Lanne's (2013) above-mentioned finding regarding realized-volatility estimators, they suggested that such estimators are more accurate for individual stocks if they do not incorporate overnight information. Some recent studies have also shown that overnight information has no significant effect on volatility prediction, and that it may even result in inferior volatility-prediction performance, as compared to estimators that do not incorporate overnight information (Dhaene and Wu, 2020; Lyócsa and Todorova, 2020).

At present, some Asia-Pacific stock markets take lunch breaks, including in China, Indonesia, Japan, Malaysia, Singapore and Thailand. This in effect creates separate morning and afternoon sessions on each trading day. However, during lunch breaks, information transmission continues, causing fluctuations, especially in opening prices in afternoon sessions. Wang et al. (2015) and Zhu et al. (2017) both found that additional leverage effects captured by negative lunch break returns and negative overnight returns played significant roles in forecasting volatility in China's stock and futures markets, respectively.

The development of China's stock market began in 1990 and has been characterized by sharp and frequent fluctuations of stock prices. Another of its characteristics has been that bull markets tend to be shorter, and bear markets longer than in other stock markets around the world. This has been ascribed to China's stock market being young and highly speculative (Wang et al., 2015). Given such characteristics, understanding the impact of information transmission on stock-price fluctuation is especially important to investors and policymakers in China.

To avoid non-trading-hours issues when making volatility calculations, Takaishi et al. (2012) separately calculated two realized volatilities (RVs) for the Tokyo Stock Exchange (TSE), one for the morning session and the other for the afternoon session and tested their statistical properties. Jayawardena et al. (2020) found that bridging the TSE's lunch break with information from adjacent markets, i.e., Australia, Hong Kong and China, led to improvements in volatility forecasting. Inspired by Takaishi et al. and Jayawardena et al.'s work, this study incorporates the information-content variables of trading and non-trading periods – including trading volume, overnight return, and lunch break return – into a heterogeneous autoregressive RV (HAR-RV) model. Then, we explore the impacts of the information content of trading and non-trading periods on volatility and attempt to achieve the improvements in the forecasting for the realized range-based volatility (RRV) of the China Securities Index 300 (CSI 300).

Additionally, due to the above-mentioned lunch breaks in China's stock market, we calculate separate RRVs for the morning and afternoon sessions of the CSI 300 and examine their relative usefulness in forecasting the RRV of the next trading day. The HAR-RV model, first proposed by Corsi (2009), is a simple approximate long-memory dynamics model of RV, and has achieved excellent forecasting performance, along with valuable economic implications. Our work extends it into an HAR-RRV model.

RRV, as proposed by Christensen and Podolskij (2007), has proven more accurate than traditional measures of RV. More recently, Christensen and Podolskij (2012) proposed realized range-based multi-power variation (RMV) and showed how it could be used both to estimate ex-post quadratic return variation and to make jump-robust inferences about integrated variance. However, the same authors suggested that realized range-based tri-power variance (RTV) was not only a more robust estimator of integrated variance than RMV, but also the most efficient estimator so far devised. Thus, we adopted RTV as the regressor in the HAR-RRV model.

This paper makes the following contributions. First, it examines whether information- content variables from trading and non-trading periods, including trading volume, overnight return, and lunch break return, have an important influence in the volatility of asset returns of China's stock market. Second, it differs from Jayawardena et al.'s (2020), which bridged non-trading periods by using intraday data from related markets and instead incorporates the RRVs of two trading sessions separated by a lunch break into the novel HAR-RRV model. To the best of our knowledge, our paper is the first to incorporate morning- and afternoon-session RRVs into a HAR model aimed at improving the forecasting of volatility in China's stock market.

In light of that primary aim, based on in-sample estimation results, we adjust and assess out-of-sample forecasting accuracy and performance, with the subsidiary aim of arriving at a forecasting model that is not only simpler, but also more accurate. The success of that effort allows us to demonstrate that, in the case of the CSI 300, the RRVs of morning sessions do play a significant role in the RRVs of the subsequent trading days.

Numerous studies have explored overnight or lunch break information content's impact on RV,[1] but little attention has thus far been paid to the parallel impact of the existence of morning and afternoon sessions that lunch breaks imply. This paper therefore complements the literature on the impacts of information from both trading and non-trading periods on RRV. Moreover, our empirical results do not support the SIA hypothesis in the case of China's stock markets. They do, however, demonstrate that overnight and lunch break returns do have a significant impact on future RRV. Moreover, having tested separate models for forecasting short-term and long-term periods of volatility, we concluded that a single model could be used to perform more accurate out-of-sample forecasting at all forecasting time horizons than our original HAR-RRV model.

[1]See, for example, Ahoniemi and Lanne (2013); Todorova and Souček (2014); Wang et al. (2015); Jayawardena et al. (2016); Zhu et al. (2017); Dhaene and Wu (2020); Jayawardena et al. (2020); and Lyócsa and Todorova (2020).

The remainder of this paper is organized as follows. Our data sources and methodology are described in Section 2, along with the specifications of the HAR-RRV and modified models. Section 3 presents our empirical findings, and the final section, our concluding remarks.

2. DATA AND METHODOLOGY

2.1. Data

There are two securities exchanges on the mainland China, the Shanghai Securities Exchange (SHSE) and the Shenzhen Securities Exchange (SZSE). They officially open at 9:30 a.m. and close at 3:00 p.m., with a lunch break from 11:30 a.m. to 1:00 p.m., giving a total of four trading hours per working day. The CSI 300, consisting of the country's 300 largest and most liquid A-share stocks, was created by China Securities Index Co., Ltd. in 2005 to serve as a proxy for the broader Chinese economy.

The CSI 300's intraday data, daily trading volume, overnight returns, and lunch break returns were collected from the iFinD database. Our empirical data for the sample period from April 11, 2005, to December 31, 2019, thereby provides 48 data periods of 5-minute duration within a single trading day, for a total of 3584 trading days. We felt that the CSI300 was a good case for exploring the impact of information transmission on stock markets, because of the more noise trader in Chinese stock markets.

2.2. Realized Volatility Measures

RV, firstly proposed by Andersen and Bollerslev (1998), is an assessment of variation in returns for an asset that uses high-frequency intraday data within a defined period. It can be divided into two components, a continuous-volatility component and a jump component, both of which influence stock prices. Prior studies have suggested that RVs' jump components would be of little help in predicting future RVs (Andersen et al., 2007). In response, Christensen and Podolskij (2007) replaced each squared intra-day return with the high-low range for the same period, to create a novel and more efficient estimator, RRV. Based on their results, they argued that intraday range-based estimation of volatility was a more accurate and efficient method of predicting volatility.

In this paper, we use intraday data from the CSI 300 to measure daily RRV and predict future RRV. Following Christensen and Podolskij (2006, 2007), the basis for our analysis is an equidistant grid, $t_j = j/n$, $j=1,\ldots,n$, where n is the sampling frequency. The intraday range is computed as

$$S_{t,j} = \text{Max}\,(p_h - p_l),\ \text{for } j = 1,2,\ldots, n \tag{1}$$

where p_h (p_l) is the highest (lowest) price during a given 5-minute period. Thus, we have a total of 48 observations of p per trading day.

Following the measures of realized range-based estimation of integrated variance used by Christensen and Podolskij (2006, 2012), we RRV and RTV as

$$RRV_t^b = \frac{1}{\lambda_2}\sum_{j=1}^{n}(s_{t,j})^2 \tag{2}$$

$$RTV_t = \frac{n}{n-2}\left(\frac{1}{\lambda_{2/3}}\right)^3 \sum_{j=3}^{n}(s_{t,j})^{2/3}(s_{t,j-1})^{2/3}\,(s_{t,j-2})^{2/3} \tag{3}$$

where $\lambda_r = 4/\sqrt{\pi}(1-4/2^r)2^{r/2}\Gamma((r+1)/2)\varsigma(r-1)$. The superscript b appearing in the formula for RRV indicates that it will be downward biased in the presence of jumps (Christensen and Podolskij, 2006).

Christensen and Podolskij (2012) proposed that RTV provides an alternative means of drawing inferences from the continuous path component of the quadratic variation. They suggested that the RTV extension is more efficient than its bi-power equivalent. Accordingly, we consider the purely range-based estimator that is consistent for the quadratic variation of the jump diffusion semi-martingale to be:

$$RRV_t \equiv \lambda_2 RRV_t^b + (1-\lambda_2)RTV_t \xrightarrow{p} \int_0^t \sigma^2(t)dt + \sum_{0<t_j<t}\kappa^2(t) \tag{4}$$

2.3. Volatility Modelling

The HAR model, proposed by Corsi (2009), holds that agents with different time horizons perceive, react to, and cause different types of volatility components. It also successfully captures the long-memory feature of realized variance for various forecasting horizons. Moreover, Christensen and Podolskij (2006) suggested that RRV is more efficient than RV. Thus, this paper uses an efficient range-based estimator to describe dynamic volatility in the HAR model. Meanwhile, we adopt RTV as the regressor for prediction of RRV. Our HAR-RRV model is therefore specified as

$$RRV_{t,t+H} = \beta_0 + \beta_D RTV_{t-1,t} + \beta_W RTV_{t-5,t} + \beta_M RTV_{t-20,t} + \varepsilon_{t,t+H} \quad (5)$$

where H denotes the prediction horizons. The measures of multi-period realized volatility, comprising $RTV_{t-5,t}$ and $RTV_{t-20,t}$, each represent the arithmetic mean of daily volatilities over the relevant time period. We focus on the prediction of future realized volatility using 1-day and 1- to 4-week horizons (i.e., H=1, 5, 10, 15 and 20).

2.4. Modified HAR Models

Innovation at the opening of the markets reflects the accumulation of information absorbed during market closure periods and is bound to have an important influence on the market during the trading day. Gallo (2001) suggested that overnight returns have explanatory power over the squared residuals of intraday returns. More recent studies have further suggested that overnight information has predictive power over, and improves the forecasting performance of, future volatility.[2]

After the opening of the market, investors' receipt of new information and resulting investment decisions will be reflected in asset prices through trading volumes. The SIA hypothesis holds that the lagged trading volume has not been completely assimilated into the prior change in price. Thus, the main implication of the SIA hypothesis is that sequential reactions to information make asset-price volatility potentially predictable via trading volume. As pointed out in previous studies, trading volume can be regarded as a proxy measure for the unobservable flow of information into the market (Andersen, 1996). More recently, Todorova and Souček (2014) concluded that trading volume was an important factor in the volatility-volume relationship.

It is also important to note that some securities exchanges in the Asia-Pacific region with lunch breaks for 1-2 hours, and thus also undergo a non-trading period of information transmission during the day. Wang et al. (2015) and Zhu et al. (2017) both found that lunch break returns had a large long-term impact on the volatility of stock and futures markets. More recently, Jayawardena et al. (2020) demonstrated that taking account of information from neighbouring markets during the TSE's lunch break led to improvements in TSE volatility forecasting.

Our primary objective is to explore the impact on RV of information transmission during periods of market opening and closure, as a means of achieving overall improvement in volatility forecasting. To this end, we propose a refined model that could improve the predictability of future realized volatility. After incorporating overnight returns, lagged trading volume, and lunch break returns into the HAR-RRV model, as described above, we further propose the following specification, to obtain our HAR-RRV-OVL model:

$$\begin{aligned} RRV_{t,t+H} = {} & \beta_0 + \beta_D RTV_{t-1,t} + \beta_W RTV_{t-5,t} + \beta_M RTV_{t-20,t} + \beta_{OD} ONR_{t-1,t} \\ & + \beta_{VD} VOL_{t-1,t} + \beta_{LD} LBR_{t-1,t} + \varepsilon_{t,t+H} \end{aligned} \quad (6)$$

where *ONR* defines the absolute value of overnight return; *VOL* defines the logarithm trading volume; *LBR* defines the positive lunch break return, that is equal to max ($LBR_{t-1,t}$, 0), which measures the leverage effects for news arrival.

Additionally, due to lunch break effects, Takaishi et al. (2012) calculated two RVs separately, one for the morning session and the other for the afternoon session of the TSE. They examined the statistical properties of both by investigating their respective standardized returns. Similarly, Jayawardena et al. (2020) decomposed total realized power variation (RPV) into morning and afternoon sessions and examined the impact on the predictability of total RPV that resulted from bridging the TSE's overnight period using global market information. Inspired by both these articles, we decomposed CSI 300 RRV into morning and afternoon sessions and incorporated these separate RRVs into our HAR-RRV model. Our original specification is modified into the HAR-RRV-MA model using the following formula,

$$\begin{aligned} RRV_{t,t+H} = {} & \beta_0 + \beta_D RTV_{t-1,t} + \beta_W RTV_{t-5,t} + \beta_M RTV_{t-20,t} + \beta_{MS} RRV^{MS}_{t-1,t} \\ & + \beta_{AS} RRV^{AS}_{t-1,t} + \varepsilon_{t,t+H} \end{aligned} \quad (7)$$

where RRV^{MS} defines the RRV of morning session; RRV^{AS} defines the RRV of afternoon session.

3. EMPIRICAL RESULTS

3.1. In-Sample Results

Table 1 presents the descriptive statistics of RRV levels for the CSI 300, as well as its overnight returns, trading volumes, lunch break returns, and the respective RRVs of its morning and afternoon sessions. Notably, the mean and standard values of RTV are smaller than those of RRV. This tends to confirm that, as pointed out by Christensen and Podolskij (2012), the range-based tri-power extension is more efficient than the bi-power one. In addition, the LB statistics for ONR, VOL, LBR, RRV^{MS},

[2] See, for example, Ahoniemi and Lanne (2013); Todorova and Souček (2014); Wang et al. (2015); Jayawardena et al. (2016); Zhu et al. (2017); Jayawardena et al. (2020).

and RRV^{AS} have a high degree of serial correlation, indicating that these variables might facilitate the prediction of future market volatility. Moreover, the mean and standard values of RRV^{MS} are higher than RRV^{AS}, implying that fluctuation during the morning session is bigger than in the afternoon one. This finding is consistent with those of Takaishi et al. (2012) regarding the statistical characteristics of TSE stock-price fluctuations.

Table 1: Descriptive statistics, realized range-based volatility levels for the CSI 300 [a]

Variable	Statistical Parameter						
	Mean	**S.D.**	**Skew**	**Kurtosis**	**Min.**	**Max.**	**$LB_{(10)}$ [b]**
RRV	1.466	2.186	5.526	52.161	0.055	35.289	8973.160
RTV	0.724	1.143	7.120	84.059	0.044	19.775	10000.747
ONR	0.422	0.626	5.236	47.803	0.000	8.577	1956.227
VOL	25.016	0.954	-0.746	4.293	21.789	27.579	31568.741
LBR	0.012	0.038	23.058	742.200	0.000	1.428	440.212
RRV^{MS}	0.879	1.392	6.009	57.389	0.025	20.788	6890.989
RRV^{AS}	0.580	1.032	6.933	80.120	0.001	19.359	5208.651

[a] *RRV*=realized range-based variance; *RTV*=realized range-based tri-power variation; *ONR*=absolute value of the overnight return; *VOL*=logarithm trading volume; *LBR*=positive lunch break return; RRV^{MS}=*RRV* of the morning session; RRV^{AS}=*RRV* of the afternoon session; S.D.=standard deviation.

[b] The critical value of the test statistics for $LB_{(10)}$ was 18.307.

The in-sample results for the CSI 300 are provided in Table 2, with Panel A providing the results of the HAR-RRV model, and Panel B, the results of the HAR-RRV-OVL model. In Panel A, all the estimates of β_D, β_W and β_M are statistically significant at least at the 10% level, except for β_W for the 2- to 4-week prediction horizon, indicating the existence of highly persistent volatility dependence. Moreover, the adjusted R^2 ranges from 0.530 for the 1-day horizon to 0.462 for the 4-week horizon, and the mean squared error (MSE) ranging from 2.256 to 1.288 across the five prediction horizons.

Table 2: In-sample estimates for the CSI 300 using the HAR-RRV and HAR-RRV-OVL models[a]

Variable	Horizon: 1 day		1 week		2 weeks		3 weeks		4 weeks	
	Coeff.[b]	S.E.	Coeff.[b]	S.E.	Coeff.[b]	S.E.	Coeff.[b]	S.E.	Coeff.[b]	S.E.
Panel A: *HAR-RRV* Model ($RRV_{t,t+H}=\beta_0+\beta_D RTV_{t-1,t}+\beta_W RTV_{t-5,t}+\beta_M RTV_{t-20,t}+\varepsilon_{t,t+H}$)										
β_0	0.274***	0.059	0.376***	0.084	0.471***	0.094	0.536***	0.096	0.580***	0.095
β_D	0.755***	0.107	0.651***	0.092	0.480***	0.093	0.341***	0.077	0.299***	0.068
β_W	0.656***	0.180	0.346**	0.148	0.215	0.181	0.263	0.204	0.279	0.210
β_M	0.236*	0.135	0.508***	0.168	0.680***	0.202	0.681***	0.229	0.645***	0.218
Adj_R^2	0.530		0.564		0.502		0.473		0.462	
MSE	2.256		1.468		1.435		1.359		1.288	
Panel B: *HAR-RRV-OVL* Model ($RRV_{t,t+H}=\beta_0+\beta_D RTV_{t-1,t}+\beta_W RTV_{t-5,t}+\beta_M RTV_{t-20,t}+\beta_{OD} ONR_{t-1,t}+\beta_{VD} VOL_{t-1,t}+\beta_{LD} LBR_{t-1,t}+\varepsilon_{t,t+H}$)										
β_0	1.032	0.721	-0.325	0.121	-1.475	1.487	-2.052	1.708	-2.490	1.791
β_D	0.745***	0.101	0.642***	0.095	0.460***	0.097	0.317***	0.080	0.273***	0.069
β_W	0.593***	0.178	0.312**	0.155	0.183	0.184	0.237	0.203	0.256	0.209
β_M	0.227*	0.127	0.493***	0.163	0.651***	0.195	0.645***	0.221	0.605***	0.210
β_{OD}	0.365***	0.096	0.194***	0.085	0.175***	0.066	0.140**	0.058	0.125**	0.058
β_{VD}	-0.033	0.029	0.027	0.045	0.077	0.060	0.103	0.069	0.123	0.072
β_{LD}	-2.167***	0.626	-1.269	0.833	0.180	1.006	1.175	1.108	1.349	1.173
Adj_R^2	0.539		0.568		0.506		0.480		0.470	
MSE	2.209		1.453		1.420		1.342		1.269	

[a] The table presents adjusted R^2 and mean squared error (MSE) for 1-day and 1- to 4-week in-sample predictions of *RRV* for the CSI 300. The S.E. values are based upon Newey-West HAC standard errors. The dependent variables for the HAR-RRV and HAR-RRV-OVL models for all horizons are standardized realized variance: $RRV_{t,t+H}/H$.

[b] *** indicates significance at the 1% level; ** indicates significance at the 5% level; and * indicates significance at the 10% level.

Similarly, as shown in Panel B of Table 2, highly persistent volatility dependence is also discernible in the HAR-RRV-OVL model. In this table, we focus on the influence of three variables on the volatility forecasting: ONR, VOL, and LBR. Firstly, it is worth noting that, in the HAR-RRV-OVL model, the estimates for β_{OD} were positive and statistically significant for all prediction horizons at least at the 5% level. Because the ONR represents the absolute value of overnight returns, the empirical results show that when important news arrives during the night, it exacerbates volatility, irrespective of whether it is good news or bad.

Additionally, estimates of β_{VD} for all prediction horizons in the HAR-RRV-OVL model were found to be insignificant at conventional levels of significance, indicating that the impact on volatility of information transmission during trading periods was not particularly important. Our empirical results from this model also do not support the SIA hypothesis, in the case of the CSI 300.

Turning to estimates of β_{LD} in the HAR-RRV-OVL model, it is interesting to note that only the β_{LD} for the 1-day prediction horizon was found to be statistically significant (and negative) at the 1% level. The LBR=max ($LBR_{t-1,t}$, 0), which represents positive lunch break returns. These empirical results reveal that good news during the lunch break mitigates volatility at shorter prediction horizons, implying a significant role of additional leverage effects captured by positive lunch break returns. But on the other hand, good news received during lunch break periods had little impact on volatility at the longer prediction horizons. These empirical findings are inconsistent with those of Wang et al. (2015) and Zhu et al. (2017), who claimed a significant role in volatility forecasting for additional leverage effects captured by negative lunch break returns and negative overnight returns.

Lastly, it should be pointed out that the explanatory and predictive abilities of our HAR-RRV-OVL model were better than those of our HAR-RRV model, as confirmed by the adjusted R^2 and MSE terms, which were higher (lower) across all prediction horizons in the HAR-RRV-OVL model.

Table 3 reports in-sample estimation results for the CSI 300 using our HAR-RRV-MA model. Here, it should be noted that the estimates for β_{MS} for the 3-week and 4-week prediction horizons using this model were positive and statistically significant at the 10% and 5% levels, respectively. This indicated that morning-session RRV had a positive impact on volatility at the longer prediction horizons. It is also interesting to note that the estimate for β_{AS} in the HAR-RRV-MA model was positive and statistically significant at the 5% level at the 1-day prediction horizon only. This differs sharply from our findings regarding β_{MS} and implies that afternoon-session RRV drives up volatility only at the shortest prediction horizon. However, adjusted R^2 (MSE) was also higher (lower) across all prediction horizons in the HAR-RRV-MA model than in the HAR-RRV model. In short, the RRVs of trading sessions, whether in the morning or the afternoon, indeed play important roles in volatility for the CSI 300, that is novel findings of this study, which also is a part of little pay attention in the prior literature.

Table 3: In-sample estimates for the CSI 300 using the HAR-RRV-MA models [a]

Variable	Horizon: 1 day		1 week		2 weeks		3 weeks		4 weeks	
	Coeff. [b]	S.E.	Coeff. [b]	S.E.	Coeff. [b]	S.E.	Coeff. [b]	S.E.	Coeff. [b]	S.E.
HAR-RRV-MA Model ($RRV_{t,t+H}=\beta_0+\beta_D RTV_{t-1,t}+\beta_W RTV_{t-5,t}+\beta_M RTV_{t-20,t}+\beta_{MS} RRV^{MS}_{t-1,t}+\beta_{AS} RRV^{AS}_{t-1,t}+\varepsilon_{t,t+H}$)										
β_0	0.245***	0.053	0.354***	0.077	0.452***	0.088	0.516***	0.090	0.556***	0.089
β_D	0.245	0.229	0.363	0.290	0.246	0.298	0.065	0.249	-0.060	0.201
β_W	0.679***	0.188	0.339**	0.143	0.209	0.175	0.260	0.197	0.285	0.201
β_M	0.226*	0.125	0.505***	0.161	0.678***	0.197	0.678***	0.222	0.640***	0.208
β_{MS}	0.197	0.135	0.181	0.126	0.147	0.101	0.159*	0.086	0.177**	0.076
β_{AS}	0.371**	0.162	0.137	0.165	0.111	0.175	0.147	0.152	0.222	0.137
Adj_R^2	0.537		0.569		0.505		0.478		0.469	
MSE	2.222		1.451		1.424		1.345		1.269	

[a] The table presents the adjusted R^2 and mean squared error (MSE) for 1-day and 1- to 4-week in-sample predictions of *RRV* for the CSI 300. The S.E. values are based upon Newey-West HAC standard errors. The dependent variables for the HAR-RRV-MA model for all horizons are standardized realized variance: $RRV_{t,t+H}/H$.
[b] *** indicates significance at the 1% level; ** indicates significance at the 5% level; and * indicates significance at the 10% level.

3.2. Out-of-Sample Results

Although the two modified models discussed above both have lower MSE values than the HAR-RRV base model, this does not necessarily mean that the modified models will significantly outperform the base model in out-of-sample forecasting. To explore whether this was the case, we conducted a modified Diebold-Mariano (DM) test, as proposed by Harvey et al. (1997). The in-sample period ran from April 11, 2005, to December 28, 2018, whilst the out-of-sample period ran from January 2, 2019 to December 31, 2019. The out-of-sample forecasting method used in this paper is the rolling-window approach, in which the sampling window is expressed relative to the forecasting date and automatically shifts forward with the passage of time. Table 4 presents the out-of-sample results of the HAR-RRV and modified models, with Panel A reporting their MSEs, and Panel B, the *t*-statistics of the modified DM test. Panel A reveals that MSE was lower in the HAR-RRV-OVL model than in the benchmark model only at the 1-day forecasting horizon. However, in the case of the HAR-RRV-MA model, MSE results were all lower than in the benchmark model, irrespective of forecasting-horizon length.

Table 4: Comparison of the Accuracy of the Three Initial Models' out-of-sample CSI 300 Forecasting Performance

Horizon	HAR-RRV	HAR-RRV-OVL	HAR-RRV-MA
Panel A: *HAR-RRV Model* [a]		MSE	
1 day	0.253	0.240	0.248
1 week	0.198	0.207	0.189
2 weeks	0.242	0.275	0.234
3 weeks	0.272	0.324	0.263
4 weeks	0.291	0.364	0.282
Panel B: Modified DM test[b]		*t-stat.*[c]	
1 day	—	1.812 **	0.337
1 week	—	-2.014	2.030 **
2 weeks	—	-1.922	2.077 **
3 weeks	—	-1.909	2.045 **
4 weeks	—	-2.022	1.963 **

[a] The table presents the out-of-sample forecasts for the CSI 300 using 5-minute intraday data covering the period from January 2, 2019 to December 31, 2019, with data from April 11, 2005 to December 28, 2018 being used to estimate the HAR-RRV model parameters. The dependent variables for all models are standardized realized variance: $RRV_{t,t+H}/H$, and the table entries represent the mean standard error (MSE) for the out-of-sample predictions based upon 1-day and 1- to 4-week out-of-sample *RRV* prediction horizons. The different columns refer to the different HAR-RRV-type models.
[b] Modified Diebold-Mariano (DM) test *p*-values are presented for 1-day and 1- to 4-week out-of-sample *RRV* predictions. The benchmark model for all horizons is the HAR-RRV model.
[c] *** indicates significance at the 1% level; ** indicates significance at the 5% level; and * indicates significance at the 10% level.

In the modified DM test (Panel B of Table 4), the HAR-RRV model was taken as the benchmark for all forecasting horizons. If *t*-statistics is positive (negative), it indicates that the examined model is more (less) accurate than the benchmark model. The Panel B results show that the HAR-RRV-OVL model performed significantly better than the benchmark model, at the 5% level, only at the 1-day forecasting horizon. Additionally, except for 1-day forecasting horizon, the HAR-RRV-MA model performed significantly better than the benchmark, again at the 5% level, implying that this model indeed improved upon overall forecasting performance of future volatility in the CSI 300. Based on the empirical results of this paper, we recommend that, to maximize forecasting accuracy, short-term periods of volatility should be forecast using the HAR-RRV-OVL model, and long-term ones, using the HAR-RRV-MA model.

3.3. More Accurate and Simplified Volatility-Forecasting Model

Next, we attempted to simplify our modeling while also improving its volatility-prediction ability. To that end, based on our in-sample estimation results, we adjusted the independent variables and re-examined its volatility forecasting for out-of-sample results. Specifically, the variables that were found to be statistically significant in the case of the in-sample results, i.e., ONR, LBR, RRV^{MS} and RRV^{AS}, were incorporated into a simplified version of the HAR-RRV model called the HAR-RRV-X model (with "X" referring to any one of ONR, LBR, RRV^{MS} or RRV^{AS}), specified as:

$$RRV_{t,t+H} = \beta_0 + \beta_D RTV_{t-1,t} + \beta_W RTV_{t-5,t} + \beta_M RTV_{t-20,t} + \beta_{XD} X_{t-1,t} + \varepsilon_{t,t+H} \quad (8)$$

Table 5 presents comparisons among the HAR-RRV and four "X" models' out-of-sample forecasting accuracy, with Panel A reporting the MSE and Panel B reporting the *t*-statistics of the modified DM test. The former panel shows that the MSE results for the HAR-RRV-ONR model were lower than for the benchmark model only at the 1-day and 1-week forecasting horizons. This was very different from the HAR-RRV-LBR model, whose MSE results were lower than those of the benchmark model at the 2-week, 3-week, and 4-week forecasting horizons. The MSE results of the HAR-RRV-AS model were slightly lower than those of the benchmark model at the 3-week and 4-week forecasting horizons. The HAR-RRV-MS model, meanwhile, had the lowest MSE results of any model at all forecasting horizons.

Table 5: Comparison of the accuracy of the four simplified models' out-of-sample CSI 300 forecasting performance against that of the benchmark model

Horizon	*HAR-RRV*	*HAR-RRV-ONR*	*HAR-RRV-LBR*	*HAR-RRV-MS*	*HAR-RRV-AS*
Panel A: *HAR-RRV Model*[a]			MSE		
1 day	0.253	0.249	0.257	0.246	0.254
1 week	0.198	0.196	0.201	0.184	0.198
2 weeks	0.242	0.243	0.240	0.228	0.242
3 weeks	0.272	0.273	0.268	0.254	0.269
4 weeks	0.291	0.293	0.285	0.268	0.287
Panel B: Modified DM test[b]			*t-stat.*[c]		
1 day	–	0.586	-2.066	1.803**	-0.032
1 week	–	1.899**	-5.476	3.120***	0.523
2 weeks	–	-1.540	4.211***	2.677***	0.953
3 weeks	–	-1.246	4.459***	3.819***	1.082
4 weeks	–	-1.084	4.286***	3.701***	1.104

[a] The table presents the out-of-sample forecasts for the CSI 300 using 5-minute intraday data covering the period from January 2, 2019 to December 31, 2019, with data from April 11, 2005, to December 28, 2018 being used to estimate the HAR-RRV model parameters. The dependent variables for all models are standardized realized variance: $RRV_{t,t+H}/H$, and the table entries represent the mean standard error (MSE) for the out-of-sample predictions based upon 1-day and 1- to 4-week out-of-sample *RRV* prediction horizons. The different columns refer to the different HAR-RRV-type models.
[b] Modified Diebold-Mariano (DM) test *p*-values are presented for 1-day and 1- to 4-week out-of-sample *RRV* predictions. The benchmark model for all horizons is the HAR-RRV model.
[c] *** indicates significance at the 1% level; ** indicates significance at the 5% level; and * indicates significance at the 10% level.

Turning now to Panel B, the HAR-RRV-ONR model's forecasting performance was significantly better than the benchmark model (at the 5% level) only at the 1-week horizon. The HAR-RRV-LBR model, on the other hand, significantly outperformed the benchmark model (at the 1% level) at the 2-week, 3-week, and 4-week forecasting horizons; and the HAR-RRV-MS model significantly outperformed the benchmark (at the 5% level or better) at all forecasting horizons. However, the HAR-RRV-AS model insignificantly outperformed the benchmark model at all forecasting horizons.

This study has demonstrated that morning-session RRV does play a significant role in RRV for the next trading day of the CSI 300. This suggests that the information content of morning sessions is more important than that of afternoon ones. A possible explanation is that information that arrives during the night is accorded more importance by traders than information that arrives during the day, and this is also reflected in more intense levels of morning-session RRV (see Table 1). In short, the HAR-RRV-MS model not only was found to be more accurate than the HAR-RRV base model, but also it was a simplified volatility-forecasting model.

4. CONCLUSIONS

Focusing on the RRV impacts of 1) information during trading and non-trading periods, and 2) differences between morning and afternoon sessions, this paper complements prior literature that has linked the RRVs of half-day trading sessions separated by formal lunch breaks to overall RRV and incorporated its findings into improved HAR models for forecasting volatility in the CSI 300.

This study has three key findings. First, it demonstrated that the SIA hypothesis is not supported in the CSI 300; instead, absolute overnight returns and positive lunch break returns were both found to play significant roles in volatility, and therefore potentially in the forecasting of volatility. These results hold for both in-sample and out-of-sample forecasting. Second, its empirical results imply that the respective RRVs of morning and afternoon sessions have important influence over longer-term volatility in the CSI 300 – implying that our HAR-RRV-MA model achieved our aim of improving the forecasting of future volatility at longer forecasting horizons (i.e., 1 week or more), while our HAR-RRV-OVL model achieved our parallel aim with regard to shorter ones (i.e., less than 1 week). Thus, thirdly, to avoid the need for using two different models for different forecasting horizons, this study proposed a simplified volatility-forecasting model that, in the event, proved the most accurate. Specifically, the HAR-RRV-MS model significantly outperformed the benchmark HAR-RRV model in terms of forecasting accuracy in the out-of-sample results at all forecasting horizons. In sum, our findings indicate that the HAR-RRV-MS model can improve the accuracy of predictions of RRV.

REFERENCES

Ahoniemi, Katja, and Markku, Lanne. (2013). Overnight stock returns and realized volatility. International Journal of Forecasting, 29(4), 592-604.

Andersen, Torben, G. (1996). Return volatility and trading volume: an information flow interpretation of stochastic volatility. Journal of Finance, 51(1), 169-204.

Andersen, Torben, G., and Tom, Bollerslev. (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. International Economic Review, 39(4), 885-905.

Andersen, Torben, G., Tim, Bollerslev, and Francis, X. Diebold. (2007). Roughing it up: including jump components in the measurement, modeling and forecasting of return volatility. Review of Economics and Statistics, 89(4), 701-720.

Christensens, Kim, and Mark, Podolskij. (2006). Range-based Estimation of Quadratic Variation. Working paper, Aarhus School of Business.

Christensens, Kim, and Mark, Podolskij. (2007). Realized range-based estimation of integrated variance. Journal of Econometrics, 141(2), 323-349.

Christensens, Kim, and Mark, Podolskij. (2012). Asymptotic theory of range-based multipower variation. Journal of Financial Econometrics, 10(3), 417-456.

Copeland, Thomas, E. (1976). A model of asset trading under the assumption of sequential information arrival. Journal of Finance, 31(4), 1149-1168.

Corsi, Fulvio. 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics, 7(2), 174-196.

Darrat, Ali, F., Shafiqur, Rahman, and Maosen, Zhong. (2003). Intraday trading volume and return of volatility the DJIA stocks: a note. Journal of Banking and Finance, 27(10), 2035-2043.

Dhaene, Geert, and Jianbin, Wu. (2020). Incorporating overnight and intraday returns into multivariate GARCH volatility models. Journal of Econometrics, 217(2), 471-495.

Gallo, Giampiero, M. (2001). Modelling the impact of overnight surprises on intra-daily volatility. Australian Economic Papers, 40(4), 567-580.

Harvey, David, Stephen, Leybourne, and Pual, Newbold. (1997). Testing the equality of prediction mean squared errors. International Journal of Forecasting, 13(2), 281-291.

Jayawardena, Nirodha, I., Neda, Todorova, Bin, Li, and Jen-Je, Su. (2016). Forecasting stock volatility using after-hour information: evidence from the Australian Stock Exchange. Economic Modelling, 52(B), 592-608.

Jayawardena, Nirodha, I., Neda, Todorova, Bin, Li, and Jen-Je, Su. (2020). Volatility forecasting using related markets' information for the Tokyo Stock Exchange. Economic Modelling, 90(C), 143-158.

Jennings, Robert, H., Laura, T., Starks, and John, C., Fellingham. (1981). An equilibrium model of asset trading with sequential information arrival. Journal of Finance, 36(1), 143-161.

Kim, Donghoon, and Goh, Jihoon. (2024). Overnight returns, daytime reversals, and anchoring bias, Applied Economics Letters: 1–5. https://doi.org/10.1080/13504851.2024.2332578.

Lyócsa, Štefan, and Neda, Todorova. (2020). Trading and non-trading period realized market volatility: does it matter for forecasting the volatility of US stocks? International Journal of Forecasting, 36(2), 628-645.

Masyhuri, Hamidi, Fajri, Adrianto, Nanda, Nanda, Eko, Dwi, Putra, and Amer, Azlan, Abdul, Jamal. (2024). Intraday return of winners vs losers Indonesian capital market evidence. International Journal of Business and Society, 25(2): 773–788.

Takaishi, Tetsuya, Ting, Ting, Chen, and Zeyu, Zheng. (2012). Analysis of realized volatility in two trading sessions of the Japanese Stock Market. Progress of Theoretical Physics Supplement, 194(May), 43-54.

Todorova, Neda, and Michael, Souček. (2014). The impact of trading volume, number of trades and overnight returns on forecasting the daily realized range. Economic Modelling, 36(C), 332-340.

Wang, Xunxiao, Chongfeng, Wu, and Weidong, Xu. (2015). Volatility forecasting: the role of lunch-break returns, overnight returns, trading volume and leverage effects. International Journal of Forecasting, 31(3), 609-619.

Zhu, Xuehong, Hongwei, Zhang, and Meirui, Zhong. (2017). Volatility forecasting using high frequency data: the role of after-hours information and leverage effects. Resources Policy, 54(C), 58-70.