

# Forecasting The Chinese Stock Market Volatility with ETF Volatility Index

Jing Li\*

School of Economics and Management, Nanjing University of Science and Technology, Nanjing, China

\*Corresponding author: 1198684571@qq.com

**Keywords:** Chinese stock market, Realized volatility, HAR model, ETF Volatility Index, Forecasting.

**Abstract:** In this paper, we forecast the realized volatility of the Shanghai Composite Index using the heterogeneous autoregressive model for realized volatility (HAR-RV) and its various extensions. Then we take a new variable named Chinese ETF volatility index into consideration, in order to compare the predictive ability between conventional models and the corresponding extended models. Our empirical results suggest that the new variable shows a significantly positive impact on the future volatility of Chinese stock market, and the extended models generate superior out-of-sample forecasting performance than the original models based on the model confidence set (MCS) test. Additionally, various sample periods, alternative volatility estimators, and alternative evaluation methods confirm the robustness of our results.

## 1. Introduction

Given that stock market volatility is central to asset pricing, asset allocation, and risk management, it is crucial to forecast volatility more accurately. Numerous studies have documented that (intra-daily) high-frequency data are useful for forecasting volatility. High-frequency volatility models measure the so-called realized volatility (RV), a concept that was pioneered by Andersen and Bollerslev (1998) [1]. Among the subsequent realized volatility models, the heterogeneous autoregressive RV (HAR-RV) model proposed by Corsi (2009) is one of the most popular [2]. Although the specification of HAR-RV is simple, it can capture “stylized facts” in financial market volatility such as long memory and multi-scaling behavior. HAR-RV has become the standard benchmark for analyzing and forecasting financial volatility dynamics (see, e.g., Yaojie Zhang et al., 2019; Yaojie Zhang et al., 2020) [3-4].

Our motivation is straightforward. The ETF volatility are documented to have strong links with stock markets volatility and return, and ETF has predictive power (see, e.g., Jose A Gutierrez et al., 2009; Qingfu Liu and Yiuman Tse, 2017; Ivan T. Ivanov and Stephen L. Lenkey, 2018; Fangfei Zhu et al., 2019) [5-8]. Mike Buckle et al. (2018) proposes that ETFs can lead the price moves, and have adjusted prices actively to pre-market information and activities [9]. It is verified by Sanjeev Bhojraj et al. (2020) that exchange-traded funds (ETFs) play in the transfer of information across firms around earnings announcements [10]. Marie-Eve Lachance (2021) examines that exchange-traded funds' (ETFs) unusually high overnight returns are distorted by market microstructure effects. Thus, we can conclude that the impact of ETF volatility will be transmitted to the stock market volatility [11]. Therefore, an increase of ETF volatility index may lead to a corresponding increase of the volatilities for stock markets. Inspired by this, this paper will research the effect of ETF volatility index on realized volatility of stock markets.

There are, however, few studies that forecast Chinese stock market volatility using ETF volatility index. To fill this gap, in this study, we forecast the realized volatility of the Shanghai Composite Index using the HAR-RV model and its extensions. We follow a related study by Wang et al. (2016) and use four of popular HAR-RV-type models summed up by that article [12]. Additionally, we add the variable ETF volatility index, in order to compare the predictive ability between conventional models and the corresponding extended models.

Our empirical results provide several notable findings. First of all, the in-sample estimation results show that our new predictor of ETF volatility index has a significantly positive impact on the realized volatility of the Chinese stock market. Secondly, the model confidence set (MCS) test proposed by Hansen et al. (2011) shows that the extended models generate superior out-of-sample forecasting performance than the original models [13]. Thirdly, we provide several robustness tests. Particularly, we perform the Direction-of-Change (DoC) test suggest by Degiannakis and Filis (2017) to explore the directional accuracy of volatility forecasts [14]. The Pesaran and Timmermann (1992) statistics suggest that the null hypothesis of no directional accuracy is rejected at the 1% significance level for all the forecasting models [15]. More importantly, our HAR-RV-RS-I-VI and HAR-RV-SJ-I-VI models yield more correct DoC rates than the original models, supporting the superiority of the two proposed models in terms of the directional accuracy. In addition, we provide evidence that our results are robust to various sample periods, alternative volatility estimators, and alternative evaluation methods.

The remainder of the paper is organized as follows. Section 2 provides the econometric specifications. Section 3 describes our data. Section 4 presents the empirical results. Section 5 details a series of robustness checks. Finally, Section 6 concludes.

## 2. Methodology

In this section, we briefly describe several popular realized volatility models.

### 2.1. Realized volatility measure

In the pioneering work of past scholars, Andersen and Bollerslev (1998) propose using realized volatility (RV) as a proxy for integrated variance [1]. For a specific business day  $t$ , the realized volatility can be calculated as the sum of the squared intraday returns:

$$RV_t = \sum_{j=1}^M r_{t,j}^2, t = 1, 2, \dots, T \quad (1)$$

Where  $r_{t,j}$  represents the  $j$ -th intraday return on day  $t$ ,  $1/M$  is the given sampling frequency.

### 2.2. Modeling realized volatility

In recent years, the heterogeneous autoregressive realized volatility (HAR-RV) of Corsi (2009) has been the most popular RV model [2]. This model accommodates some of the stylized facts found in financial asset return volatility such as long memory and multi-scaling behavior. The HAR-RV is simple to implement, as it only contains three predictors: including lagged daily realized volatility ( $RV_{d,t}$ ), lagged weekly realized volatility ( $RV_{w,t}$ ), and lagged monthly realized volatility ( $RV_{m,t}$ ). The HAR-RV model specification can be expressed as

$$RV_{t+1} = \beta_0 + \beta_d RV_t + \beta_w RV_{w,t} + \beta_m RV_{m,t} + \varepsilon_{t+1} \quad (2)$$

where  $RV_{t-h:t-1} = (1/h)(RV_{t-h} + \dots + RV_{t-1})$ . Thus  $RV_{w,t}$  is the average RV from day  $t-4$  to day  $t$  and  $RV_{m,t}$  is the average RV from day  $t-21$  to  $t$ .

To capture the role of the ‘‘leverage effect’’ in volatility dynamics, Patton and Sheppard (2015) develop a series of models using signed realized measures [16]. The first model extends the standard HAR-RV by decomposing daily RV into two semi-variances (HAR-RV-RS-I),

$$RV_{t+1} = \beta_0 + \beta_d^+ RS_t^+ + \beta_d^- RS_t^- + \beta_w RV_{w,t} + \beta_m RV_{m,t} + \varepsilon_{t+1} \quad (3)$$

Where  $RS_t^- = \sum_{j=1}^M r_{t,j}^2 I(r_{t,j} < 0)$  and  $RS_t^+ = \sum_{j=1}^M r_{t,j}^2 I(r_{t,j} > 0)$

The second model for capturing the ‘‘leverage effect’’ contains a signed jump variation and an estimator of the variation caused by the continuous part (bi-power variation) (HAR-RV-SJ-I):

$$RV_{t+1} = \beta_0 + \beta_{j,d} SJ_t + \beta_{bv,d} BV_t + \beta_w RV_{w,t} + \beta_m RV_{m,t} + \varepsilon_{t+1} \quad (4)$$

Where  $SJ_t = RS_t^+ - RS_t^-$ ,  $BPV_t = u_1^{-2} \sum_{j=2}^M |r_{t,j-1}| |r_{t,j}|$ ,  $u_1 = (2/\pi)^{0.5} = E(|Z|)$ .

The last model for the ‘leverage effect’ decomposes the role of the positive and negative jumps, which is termed HAR-RV-SJ-II,

$$RV_{t+1} = \beta_0 + \beta_{j,d}^+ SJ_t^+ + \beta_{j,d}^- SJ_t^- + \beta_{bv,d} BV_t + \beta_w RV_{w,t} + \beta_m RV_{m,t} + \varepsilon_{t+1} \quad (5)$$

Where  $SJ_t^+ = SJ_t I(SJ_t > 0)$  and  $SJ_t^- = SJ_t I(SJ_t < 0)$ .

Thus, we have four HAR-type volatility equations to model and forecast realized volatility. Meanwhile, in order to compare the forecasting results with and without ETF Volatility Index in the model, our research will this variable to the above four classic models.

### 3. Data

To forecast the Chinese stock market volatility, we use CBOE China ETF Volatility Index (VIX) from FRED Economic Research. The Shanghai Composite Index (SSEC) serves as a representative of the Chinese stock market, the realized variance of which is obtained from the Oxford-Man Institute’s Quantitative Finance Realized Library. Because Liu et al. (2015) find little evidence that the 5-min RV is outperformed by any other measures from 400 volatility estimators for 31 different financial assets spanning five asset classes [17]. Therefore, we use 5-min RV from the SSEC index.

Our sample period is from April 18, 2011 to June 22, 2021. After matching the days on which all the considered the Chinese stock market trading days have corresponding VIX, we obtain 2392 observations for the stock market. To generate out-of-sample forecasts, we divide the whole sample period into an in-sample estimation period consisting of the first 1000 observations and an out-of-sample evaluation period consisting of the remaining 1392 observations.

Figure 1 depicts the evolution of the Chinese stock market’s realized volatility and China ETF Volatility Index. Their corresponding descriptive statistics are shown in Table 1. All of the series are right skewed except SJ and SJ-, and they all display positive kurtosis, suggesting they have non-normal distributions. The Jarque-Bera statistic rejects the null hypothesis of a normal distribution for these variables at 1% significant level, further confirming the fat-tailed distribution. Finally, the null hypothesis of a unit root is rejected at the 1% significance level based on the Augmented Dickey-Fuller (ADF) statistics. This evidence suggests that all the time series of these realized variances are stationary and thereby can be employed directly without further transformations.

Table 1. Descriptive statistics

Variables	Min	Max	Mean	Std. Dev.	Skewness	Kurtosis	Jarque-Bera	ADF
RV	0.046	41.460	1.220	2.628	8.277	92.046	868140.14	-5.04***
RV <sub>w</sub>	0.113	30.723	1.221	2.182	6.335	54.992	316102.54	-4.35***
RV <sub>m</sub>	0.160	15.147	1.221	1.831	4.485	24.232	66275.28	-3.42**
BPV	0.050	45.308	1.114	2.500	8.886	106.595	1159134.79	-4.84***
RS +	0.016	20.894	0.605	1.396	7.844	81.325	680878.21	-4.51***
RS-	0.018	27.725	0.616	1.363	9.533	127.748	1655808.26	-5.46***
SJ	-14.54	11.429	-0.011	0.840	-1.169	88.015	769418.07	-10.23***
SJ+	0.00	11.429	0.149	0.578	9.515	121.608	1503852.63	-5.12***
SJ-	-14.54	0.000	-0.160	0.569	-13.797	269.673	7294163.17	-10.42***
ETF_VI	15.09	69.280	25.688	6.754	1.732	4.678	3364.36	-5.85***

This table reports the descriptive statistics of all the used variables. The Jarque-Bera statistic is used to test the null hypothesis of normal distribution. ADF is the Augmented Dickey-Fuller statistic. All realized variances here are multiplied by 10000. The entire sample period is from April 18, 2011, to June 22, 2021. \*\*\*Indicates significance at the 1% level. \*\* Indicates significance at the 5% level. \*Indicates significance at the 10% level.

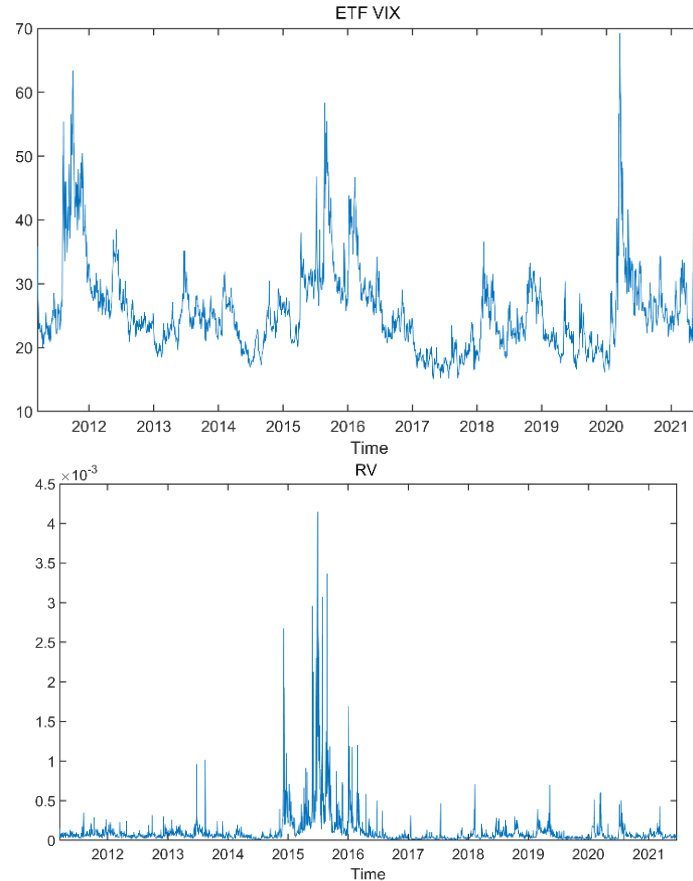


Figure 1. The realized variances of SSEC and China ETF Volatility Index.  
The sample period is from April 18, 2011 to June 22, 2021.

#### 4. Empirical results

In this section, we first give the in-sample estimation results of the eight volatility models used in study. Then, we evaluate the out-of-sample forecasting performances of the eight models with and without ETF Volatility Index.

##### 4.1. In-sample estimation results

Table 2 and table 3 show the OLS estimates of the eight volatility models over the whole sample period along with the t-statistics based on the Newey–West. From the two tables, we can see that the parameter estimates of daily, weekly, and monthly RV in each model are all significant at the 1% level, suggesting strong persistence in the realized volatility dynamics. The coefficients of leverage effect in the HAR-RV-RS-I, HAR-RV-SJ-I, HAR-RV-SJ-II and corresponding extended models added ETF volatility Index are significant, indicating that it plays an important role in the volatility process. Moreover, we examine the null hypothesis that positive and negative semi-variances have equal predictive power for realized volatility (i.e.,  $\beta_a^+ = \beta_a^-$ ) based on the benchmark model and corresponding extended model. The parameter estimates of the signed jumps,  $\beta_{j,d}^+$  is significantly positive while  $\beta_{j,d}^-$  is not significantly, indicating that only positive (“good”) jumps lead to higher future volatility.

Additionally, the regression coefficients of the Chinese ETF volatility Index are positive and significant at the 1% level in the latter four models. It indicates the index has a positive impact on the volatility of Chinese stock market. Taking the adjusted  $R^2$  into consideration, we can find that the value of the adjusted  $R^2$  is more than 68% and the value of which has increased in the four extended models. The overall model fits better.

Table 2. Estimation results of realized volatility models for the whole sample

	HAR-RV	HAR-RV-RS-I	HAR-RV-SJ-I	HAR-RV-SJ-II
$\beta_0$	-0.846*** (-6.632)	-0.623*** (-4.795)	-0.842*** (-6.611)	-1.077*** (-7.122)
$\beta_d$	0.349*** (14.033)			
$\beta_w$	0.350*** (9.563)	0.377*** (10.355)	0.385*** (10.749)	0.382*** (10.682)
$\beta_m$	0.220*** (7.496)	0.214*** (7.347)	0.218*** (7.432)	0.220*** (7.538)
$\beta_d^+$		0.279*** (14.569)		
$\beta_d^-$		0.046** (2.077)		
$\beta_{j,d}$			667.093*** (5.138)	
$\beta_{bv,d}$			0.314*** (13.063)	0.291*** (11.530)
$\beta_{j,d}^+$				1.163e+03*** (5.385)
$\beta_{j,d}^-$				192.548 (0.917)
$\beta_{VI}$				
Adj.R <sup>2</sup>	0.6804	0.6864	0.6826	0.6835

This table provides the parameter estimation results for the four realized volatility models for the whole sample period from April 18, 2011 through June 22, 2021. The numbers in the parentheses are the t-statistics. The asterisks \*, \*\* and \*\*\* denote rejections of null hypothesis at 10%, 5% and 1% significance levels, respectively.

Table 3. Estimation results of extended realized volatility models for the whole sample

	HAR-RV-VI	HAR-RV-RS-I-VI	HAR-RV-SJ-I-VI	HAR-RV-SJ-II-VI
$\beta_0$	-1.767*** (-5.986)	-1.523*** (-5.175)	-1.878*** (-6.396)	-2.107*** (-6.933)
$\beta_d$	0.345*** (13.882)			
$\beta_w$	0.343*** (9.375)	0.369*** (10.123)	0.374*** (10.448)	0.371*** (10.382)
$\beta_m$	0.200*** (6.689)	0.194*** (6.556)	0.195*** (6.547)	0.198*** (6.652)
$\beta_d^+$		0.276*** (14.404)		
$\beta_d^-$		0.047** (2.121)		
$\beta_{j,d}$			669.602*** (5.173)	
$\beta_{bv,d}$			0.312*** (13.023)	0.290*** (11.498)
$\beta_{j,d}^+$				1.163e+03*** (5.395)
$\beta_{j,d}^-$				198.723 (0.949)
$\beta_{VI}$	0.193*** (3.458)	0.189*** (3.406)	0.218*** (3.913)	0.217*** (3.902)
Adj.R <sup>2</sup>	0.6819	0.6878	0.6845	0.6854

This table provides the parameter estimation results for the four extended realized volatility models for the whole sample period from April 18, 2011 through June 22, 2021. The numbers in the parentheses are the t-statistics. The asterisks \*, \*\* and \*\*\* denote rejections of null hypothesis at 10%, 5% and 1% significance levels, respectively.

#### 4.2. Out-of-sample forecasting performance

The out-of-sample forecasting results are more important than the in-sample estimation results. Therefore, we generate the out-of-sample forecasts of the Chinese realized volatility based on a rolling estimation window. More specifically, our entire sample is divided into an in-sample portion composed of the first 1000 observations and an out-of-sample portion composed of the remaining 1392 observations. When we obtain each out-of-sample forecast, we should roll forward the estimation sample by adding one new observation and dropping the first one in the previous estimation window.

To quantitatively compare the out-of-sample performance among the forecasting models used by this paper, we apply two popular loss functions of QLIKE and MSE. In particular, Patton (2011) demonstrates that, in terms of the ranking of competing volatility forecasts, QLIKE and MSE are robust to the presence of noise in the volatility proxy [18]. MSE is the mean squared error of realized variance forecasts. Statistically, QLIKE, and MSE can be expressed as follows:

$$QLIKE = \frac{1}{q} \sum_{t=m+1}^{m+q} \left( \log(\widehat{RV}_t) + \frac{RV_t}{\widehat{RV}_t} \right) \quad (6)$$

$$MSE = \frac{1}{q} \sum_{t=m+1}^{m+q} (RV_t - \widehat{RV}_t)^2 \quad (7)$$

Where  $RV_t$  is the actual RV on day t,  $\widehat{RV}_t$  is the RV forecast based on one of the forecasting models, and m and q are the length of in-sample estimation period and out-of-sample evaluation period, respectively.

We assess the statistical significance of differences in forecasting losses using the model confidence set (MCS) proposed by Hansen et al. (2011) [13]. A MCS is a subset of models that contains the best model with a given level of confidence. Following Hansen et al. (2011), Wang et al. (2016), and Liu et al. (2018), among others, we use the confidence level of 90% [12-13,19]. This allows us to exclude a model with a p-value smaller than 0.1 from the MCS. In other words, the forecasts of this model are significantly less accurate than the models in the MCS. It is evident that a model with a larger MCS p-value shows stronger predictive ability.

Table 4 reports the out-of-sample forecasting performance of all the models, including the mean value of loss functions and the MCS p-values for the paired comparison of the HAR-RV type models and the extended models added ETF volatility Index. An impressive finding is that only two conventional models consistently appear in the MCS with the confidence level of 90%. Furthermore, the extended models always deliver the largest MCS p-value of 1. This evidence suggests that the method of adding volatility index has significantly better out-of-sample forecasting ability than the other conventional forecasting models in the prediction of the Chinese stock market volatility. This implies that the Chinese ETF volatility Index is a rather efficient variable to improve the forecast accuracy of the Chinese stock market volatility in the context of the country.

Table 4. Out-of-sample forecasting performances evaluated by loss functions based on the MCS test

Forecasting models	QLIKE		MSE	
	Loss function	P-value	Loss function	P-value
HAR-RV	0.629	<b>0.118</b>	2.061	<b>0.244</b>
HAR-RV-RS-I	0.629	0.088	2.005	<b>0.247</b>
HAR-RV-SJ-I	0.631	0.082	1.863	<b>0.258</b>
HAR-RV-SJ-II	0.630	<b>0.112</b>	1.849	<b>0.693</b>
HAR-RV-VI	0.627	<b>1.000</b>	2.025	<b>1.000</b>
HAR-RV-RS-I-VI	0.627	<b>1.000</b>	1.969	<b>1.000</b>
HAR-RV-SJ-I-VI	0.629	<b>1.000</b>	1.812	<b>1.000</b>
HAR-RV-SJ-II-VI	0.628	<b>1.000</b>	1.827	<b>1.000</b>

This table reports the out-of-sample forecasting performance of eight HAR-RV-type models including the value of two loss functions and the MCS p-values. Two loss functions we consider are QLIKE and MSE. Bold figures highlight instances in which the MCS p-value is larger than 0.1. The entire sample period containing 2392 observations spans from April 18, 2011, to June 22, 2021, while the length of in-sample period is 1000.

## 5. Robustness test

### 5.1 Various sample periods

Rossi and Inoue (2012) and Inoue, Jin, and Rossi (2017) emphasize that the arbitrary choices of different window sizes may result in quite different out-of-sample results in practical applications [20-21]. (see, e.g. Jing Liu et al., 2019; Yaojie Zhang et al., 2020) [22-23]. Therefore, the forecasting window size plays a crucial role in out-of-sample evaluation. Considering this, we additionally consider another two window sizes, where the initial in-sample estimation windows contain 800 and 1200 observations, so that the corresponding out-of-sample length is 1592 and 1192, respectively. It should be noted that both the two forecasting windows considered in this paper have a desirable trade-off between an initial in-sample estimation period that has enough observations to precisely estimate parameters and an out-of-sample period that has a relatively long length for forecast evaluation.

Table 5 reports the MCS p-values for an alternative forecasting window, in which the length of in-sample periods is 800 and 1200. Analogously, this evidence suggests that the HAR-RV extended models are more likely to be the best model for Chinese stock markets under alternative window sizes. The extended models always deliver the largest MCS p-value of 1. Although the classic HAR-RV type models illustrate good results, the method of adding volatility index has significantly better out-of-sample forecasting ability than the other conventional forecasting models. In conclusion, the out-of-sample results are robust to various sample periods.

Table 5. Out-of-sample performance for different in-sample evaluation periods

Forecasting models	QLIKE		MSE	
	Loss function	P-value	Loss function	P-value
Panel A: the length of in-sample evaluation is 800				
HAR-RV	0.808	0.451	4.989	0.546
HAR-RV-RS-I	0.809	0.328	4.867	0.466
HAR-RV-SJ-I	0.813	0.313	4.852	0.195
HAR-RV-SJ-II	0.813	0.340	5.252	0.429
HAR-RV-VI	0.807	<b>1.000</b>	4.955	<b>1.000</b>
HAR-RV-RS-I-VI	0.808	<b>1.000</b>	4.826	<b>1.000</b>
HAR-RV-SJ-I-VI	0.811	<b>1.000</b>	4.749	<b>1.000</b>
HAR-RV-SJ-II-VI	0.812	<b>1.000</b>	5.161	<b>1.000</b>
Panel B: the length of in-sample evaluation is 1200				
HAR-RV	0.397	0.120	0.326	0.837
HAR-RV-RS-I	0.398	0.090	0.323	0.837
HAR-RV-SJ-I	0.400	0.112	0.331	0.950
HAR-RV-SJ-II	0.400	0.096	0.330	0.827
HAR-RV-VI	0.395	<b>1.000</b>	0.325	<b>1.000</b>
HAR-RV-RS-I-VI	0.396	<b>1.000</b>	0.322	<b>1.000</b>
HAR-RV-SJ-I-VI	0.398	<b>1.000</b>	0.330	<b>1.000</b>
HAR-RV-SJ-II-VI	0.398	<b>1.000</b>	0.329	<b>1.000</b>

This table reports the out-of-sample forecasting performance of 8 HAR-RV-type models including the value of two loss functions and the MCS p-values. Two loss functions we consider are QLIKE and MSE. Bold figures highlight instances in which the p-value is equal to 1. The entire sample period containing 2392 observations spans from April 18, 2011, to June 22, 2021, while the length of in-sample period includes 800(Panel A) and 1200(Panel B).

## 5.2 Alternative volatility estimator

In view of the unobservable actual volatility, we further consider another prevailing volatility estimator, the realized kernel (RK) proposed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), to re-examine the out-of-sample performance for the eight above-mentioned forecasting models [24]. An appealing property of the realized kernel is that this volatility estimator is robust to market microstructure noise (Barndorff-Nielsen et al., 2008) [24]. Mathematically, the realized kernel for market  $i$  on trading day  $t$  can be defined as

$$RK_t = \sum_{h=-H}^H k\left(\frac{h}{H+1}\right) \gamma_h \quad (8)$$

were

$$\gamma_h = \sum_{j=|h|+1}^M r_{t,j} r_{t,j-|h|} \quad (9)$$

and  $k(x)$  is the Parzen kernel function, which is given by

$$k(x) = \begin{cases} 1 - 6x^2 + 6x^3 & 0 \leq x \leq 1/2 \\ 2(1-x)^3 & 1/2 \leq x \leq 1 \\ 0 & x > 1 \end{cases} \quad (10)$$

It is necessary for  $H$  to increase with the sample size in order to consistently estimate the increments of quadratic variation in the presence of noise. We follow precisely the bandwidth choice of  $H$  spelt out in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009), which we refer the reader to for more details [25]. The data of the realized kernel are also available from the Oxford-Man Institute's Realized Library.



We replace the realized volatility with the realized kernel in all the used models and then re-run these regression models to generate the realized kernel forecasts. Table 6 reports the MCS p-values when we use the realized kernel to estimate, forecast, and evaluate the Chinese stock market volatilities. We observe a robust result that the Chinese ETF volatility Index continues to show very powerful predictive ability. All the extended models always deliver the largest MCS p-value of 1 based on the loss functions QLIKE and MSE except one. In summary, the results of the MCS test are robust to alternative volatility estimators.

Table 6. MCS p-values for an alternative volatility estimator of realized kernel

Forecasting models	QLIKE		MSE	
	Loss function	P-value	Loss function	P-value
HAR-RV	0.688	0.071	2.490	0.178
HAR-RV-RS-I	0.679	0.040	2.351	0.155
HAR-RV-SJ-I	0.683	0.043	2.324	0.159
HAR-RV-SJ-II	0.683	0.047	2.498	<b>1.000</b>
HAR-RV-VI	0.685	<b>1.000</b>	2.430	<b>1.000</b>
HAR-RV-RS-I-VI	0.676	<b>1.000</b>	2.286	<b>1.000</b>
HAR-RV-SJ-I-VI	0.679	<b>1.000</b>	2.244	<b>1.000</b>
HAR-RV-SJ-II-VI	0.679	<b>1.000</b>	2.510	0.864

This table reports the out-of-sample forecasting performance of 8 HAR-RV-type models including the value of two loss functions and the MCS p-values. We use an alternative volatility estimator of the realized kernel. Two loss functions we consider are QLIKE and MSE. Bold figures highlight instances in which the p-value is equal to 1. The entire sample period containing 2392 observations spans from April 18, 2011, to June 22, 2021.

### 5.3 Alternative evaluation methods

Although the model confidence set (MCS) proposed by Hansen, Lunde, and Nason (2011) is very suitable for this study, we need employ other methods to consolidate our research results [13].

Following Degiannakis and Filis (2017), we employ the Direction-of-Change (DoC) as an additional out-of-sample evaluation criterion [14]. Degiannakis and Filis (2017) state that the DoC is central to the trading strategies of market timing and asset allocation [14]. Specifically, the DoC measures the proportion of forecasts that correctly predict the direction of the volatility movement. We let  $p_t$  be a dummy variable that takes the value of one if a model correctly predicts the direction of volatility movement on trading day  $t$ , and zero otherwise. Consequently, this dummy variable is given by

$$p_t = \begin{cases} 1 & \text{if } RV_t > RV_{t-1} \text{ and } \widehat{RV}_t > RV_{t-1} \\ 1 & \text{if } RV_t < RV_{t-1} \text{ and } \widehat{RV}_t < RV_{t-1} \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

Further, we define the DoC rate as the proportion of forecasts that correctly predict the direction of the volatility movement. Statistically, the DoC rate is equal to  $1/q \sum_{t=m+1}^{m+q} p_t$ . To explore the statistical significance of the directional accuracy, we use a nonparametric test proposed by Pesaran and Timmermann (1992) to test the null hypothesis that the DoC rate of a forecasting model of interest is less than or equal to the DoC rate of random walk against the alternative hypothesis that the DoC rate of a forecasting model of interest is larger than the DoC rate of random walk [15].

Table 7 reports the DoC results for all the forecasting models. First of all, we reject the null hypothesis of no directional accuracy at the 1% significance level for all the forecasting models, suggesting the success of the HAR-RV model as well as its extended models in the directional prediction. Second and more importantly, the two extended models (HAR-RV-RS-I-VI and HAR-RV-SJ-I-VI) yield larger DoC rates than the corresponding original HAR-RV models. Also, the confidence levels for the higher DoC rates of the extended models are greater than those of the conventional HAR-

RV models. On average, the HAR-RV-RS-I-VI model yields the largest DoC rate of 0.6334. In conclusion, the DoC results are basically consistent with the MCS results. The extended models commonly exhibit better predictive ability than the other original models from a directional accuracy perspective.

Table 7. Direction-of-Change rates

Forecasting models	DoC rate	PT statistic
HAR-RV	0.6211***	10.185
HAR-RV-RS-I	0.6326***	10.988
HAR-RV-SJ-I	0.6183***	9.913
HAR-RV-SJ-II	0.6226***	10.222
Extend models		
HAR-RV-VI	0.6175***	9.961
HAR-RV-RS-I-VI	<b>0.6334***</b>	<b>11.090</b>
HAR-RV-SJ-I-VI	<b>0.6262***</b>	<b>10.545</b>
HAR-RV-SJ-II-VI	0.6219***	10.155

This table reports the Direction-of-Change (DoC) rates and the PT statistics of Pesaran and Timmermann (1992). We consider all the forecasting models used by this paper. Bold and underlined figures highlight instances in which the DoC rate is the larger than the corresponding conventional models. Statistical significance for DoC rate is based on the p-values of the PT statistic. Note that the null hypothesis of no directional accuracy is rejected at the 1% level for all of the reported DoC rates. The entire sample period containing 2392 observations spans from April 18, 2011, to June 22, 2021, while the length of in-sample period is 1000. \*\*\*Indicates significance at the 1% level.

## 6. Conclusions

In this paper, we predict the Chinese stock market (SSEC) realized volatility based on the HAR-RV framework, which includes four popular HAR-RV-type models and four extended models by adding the variable Chinese ETF volatility index. The in-sample results suggest that the new variable shows a significantly positive impact on the future volatility of Chinese stock market.

Using the MCS test to evaluate out-of-sample forecasting performance, we find that the four extend models exhibit significantly better out-of-sample forecasting performance than the four conventional HAR-RV type models. In addition to this, our results are robust to various sample periods, alternative volatility estimators (i.e., realized variance and realized kernel), and alternative evaluation methods. In terms of directional accuracy, the HAR-RV-RS-I-VI and HAR-RV-SJ-I-VI models generate higher DoC rates, suggesting superior predictive ability relative to the original models. In conclusion, the extended models performs better than classic HAR-RV volatility forecasting models.

## References

- [1] Andersen T G, Bollerslev T. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts [J]. *International Economic Review*,1998,39: 885– 905.
- [2] Corsi F. A simple approximate long-memory model of realized volatility [J]. *Journal of Financial Economic*, 2009,7: 174–196.
- [3] Yaojie Zhang, Feng Ma, Yu Wei. Out-of-sample prediction of the oil futures market volatility: A comparison of new and traditional combination approaches [J]. *Energy Economics*, 2019,81: 1109-1120.
- [4] Yaojie Zhang, Likun Lei, Yu Wei. Forecasting the Chinese stock market volatility with international market volatilities: The role of regime switching [J]. *The North American Journal of Economics and Finance*, 2020,52: 101145.
- [5] Jose A Gutierrez, Valeria Martinez, Yiuman Tse. Where does return and volatility come from? The case of Asian ETFs [J]. *International Review of Economics & Finance*, 2009,18: 671-679.
- [6] Qingfu Liu and Yiuman Tse. Overnight returns of stock indexes: Evidence from ETFs and futures [J]. *International Review of Economics & Finance*, 2017,48: 440-451.
- [7] Ivan T Ivanov, Stephen L Lenkey. Do leveraged ETFs really amplify late-day returns and volatility? [J]. *Journal of Financial Markets*,2018,41: 36-56.
- [8] Fangfei Zhu, Xingguo Luo, Xuejun Jin. Predicting the volatility of the iShares China Large-Cap ETF: What is the role of the SSE 50 ETF? [J]. *Pacific-Basin Finance Journal*,2019,57: 101192.
- [9] Mike Buckle, Jing Chen, Qian Guo, Chen Tong. Do ETFs lead the price moves? Evidence from the major US markets. *International Review of Financial Analysis*,2018,58: 91-103.
- [10] anjeev Bhojraj, Partha Mohanram, Suning Zhang. ETFs and information transfer across firms [J]. *Journal of Accounting and Economics*,2020,70: 101336.
- [11] Marie-Eve Lachance. ETFs' high overnight returns: The early liquidity provider gets the worm [J]. *Journal of Financial Markets*,2021,52: 100563.
- [12] Wang Y, Ma F, Wei Y, Wu C. Forecasting realized volatility in a changing world: a dynamic model averaging approach [J]. *Journal of Banking & Finance*, 2016, 64: 136–149.
- [13] Hansen P R, Lunde A, Nason J M. The model confidence set [J]. *Econometrica*, 2011,79: 453–497.
- [14] Degiannakis S, Filis G. Forecasting oil price realized volatility using information channels from other asset classes [J]. *Journal of International Money and Finance*, 2017,76: 28–49.
- [15] Pesaran M H, Timmermann A. A simple nonparametric test of predictive performance [J]. *Journal of Business & Economic Statistics*, 1992,10: 461–465.
- [16] Patton A J, Sheppard K. Good volatility, bad volatility: signed jumps and the persistence of volatility [J]. *The Review of Economics and Statistics*, 2015, 97: 683–697.
- [17] Liu L Y, Patton A J, Sheppard K. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes [J]. *Journal of Econometrics*, 2015,187: 293–311.
- [18] Patton A J. Volatility forecast comparison using imperfect volatility proxies [J]. *Journal of Econometrics*, 2011,160: 246–256.
- [19] Liu J, Ma F, Yang K, Zhang Y. Forecasting the oil futures price volatility: large jumps and small jumps [J]. *Energy Economics*, 2018,72: 321–330.
- [20] Rossi B, Inoue A. Out-of-sample forecast tests robust to the choice of window size [J]. *Journal of Business & Economic Statistics*,2012, 30: 432–453.

- [21] Inoue A, Jin L, Rossi B. Rolling window selection for out-of-sample forecasting with time-varying parameters [J]. *Journal of Econometrics*,2017, 196: 55–67.
- [22] Jing Liu, Feng Ma, Yaojie Zhang. Forecasting the Chinese stock volatility across global stock markets [J]. *Physica A: Statistical Mechanics and its Applications*, 2019, 525: 466-477.
- [23] Yaojie Zhang, Feng Ma, Yin Liao. Forecasting global equity market volatilities [J]. *International Journal of Forecasting*,2020, 36: 1454-1475.
- [24] Barndorff-Nielsen O E, Hansen P R, Lunde A, Shephard N. Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise [J]. *Econometrica*,2008,76: 1481-1536.
- [25] Barndorff-Nielsen O E, Hansen P R, Lunde A, Shephard N. Realized kernels in practice: Trades and quotes [J]. *The Econometrics Journal*,2009, 12: 1–32.