International Journal of
Information
and
Management
Sciences

International Journal of Information and Management Sciences
30 (2019), 283-304. DOI:10.6186/IJIMS.201912_30(4).0001

Comparing and Combining Realized Measure and Implied Volatility for Volatility Prediction

Shi Yafeng[1], *Yanlong Shi*[2], *Peng Xun*[3], *Zhu Nenghui*[4], *Ying Tingting*[5*] *and Yan Ju*[6]

[1]Ningbo University of Technology, [2]Zhejiang Pharmaceutical College,
[3]Guangdong Guangya High School, [4]Xiamen University of Technology,
[5]University of Nottingham Ningbo China and [6]Shanxi University of Finance and Economics

Abstract

In this study, we compare realized measure and implied volatility for forecasting of the volatility. Firstly, we employ a family of homogenous loss functions as the evaluation criteria in order to run a fair and complete forecast comparison. Our results show that, predictors based on realized measures are superior to that derived from implied volatility for both within-sample fitting and one-step-ahead forecasting, whereas the superiority of the latter type is presented in multi-step-ahead forecasting. Secondly, as a result of the comparison, a new model average approach with weights that depend on market conditions is developed to combine the information from implied volatility and realized measure for volatility forecasting. And our results show the superiority of the proposed approach in combining realized measures and implied volatility for volatility prediction.

Keywords: Volatility forecasts, forecasting competitions, combining forecasts, realized volatility, implied volatility.

1. Introduction

As non-parametric daily realized measures of volatility constructed from high-frequency data has become prevailing, a new comparison analysis based on information content of realized measure and implied volatilities from options in forecasting volatility has occurred. Using an implied volatility index (VIX) which derives from S&P 100 or 500 index options, Blair et al. [5] provide the evidence for the insignificance of incremental forecasting information in the realized measures. In contrast, Koopman et al. [13] show that models with the realized measure are more superior. From recent study of Han and Park [11] (henceforth HP) it can be concluded that, for in-sample fitting, forecasts derived from time-series models of realized measures outperform those from implied volatility. In contrast, the latter is more informative than the former for out-of-sample forecasting, in particular for multi-step-ahead forecasting. Therefore, there have mixed conclusions so far.

The aforementioned researches have predominantly focused on ranking predictors by employing certain robust loss function as forecast evaluation criteria. Using certain robust loss function as forecast evaluation criteria may provide consistent ranking of them, but it obviously can't provide a complete investigation into their forecasting performances. Recently, Patton [17] propose a family of homogenous loss functions that are robust to the presence of noise in realized volatility proxies and contain the widely used criteria as special cases. Obviously, with the aid of these loss function, a more complete investigation can be performed. Moreover, for time-series models of realized measures, there have been some new developments in current literatures; see, Patton and Sheppard [18] find that disentangling the effects of negative and positive realized semivariance can significantly improve forecasts of future volatility. These new developments may have potential to provide new evidences related to previous studies.

In addition, the previous studies comparing realized measures with implied volatility all has employed the VIX index that based on the S&P 500 or 100 index as the sample data. It is well known that the VXD and VXN indexes based on DJIA and Nasdaq 100 indices respectively are available for a long period. In order to obtain more convincing results, in addition to VIX index, the sample data of our study also includes the VXD and VXN indexes as well. Clearly, these facts warrant a fresh investigation into the competition between forecasts from implied volatility and time-series models of realized volatility.

Finally, we undertake a model average of forecasts considered in this paper. Although there are some literature devoted to combine the forecasts produced by abovementioned two approaches, the results are not always desirable. For example, Han and Park [11] had recently pointed out, it is not helpful to use all the information provided by the realized measure and the implied volatility within the framework of GARCH-X model for multi-step-ahead forecasting. However, the intuition is that, in general, the more information are utilized, the more precise forecasts will be obtained under proper model framework. And as pointed by Clemen [6], the results have been virtually unanimous: combining multiple forecasts leads to increased forecast accuracy. Therefore, looking for a proper approach to combine them is crucial in increasing forecast accuracy.

Thus, contributions of our paper are multifold: Firstly, it updates earlier research by employing a family of homogenous loss functions as the evaluation criteria, and considering state-of-the-art forecasting models of realized measures. Secondly, it focuses on new implied volatility index. Obviously, compared with previous studies that only use S&P 500 or S&P 100 index data, our study applies more sample data, and consequently, the obtained results are more convincing. Finally, it proposes a new model average approach with weights that depend on market condition to combine the forecasts.

Within the Diebold-Mariano and West test (see West [19]) framework, the forecast evaluations show that, predictors based on realized measures are superior to those derived from implied volatility for within-sample fitting and one-step-ahead forecasting, whereas the latter's superiority lies in multi-step-ahead forecasting and enhances gradually with increasing forecast horizon. Comparing with previous studies, the proposed model average approach performs the best for multi-step-ahead out-of-sample forecast, especially for 10-step-ahead and 22-step-ahead forecast.

This paper is organized as follows. In Section 2, we reinvestigate the forecasting behaviors of volatility predictors using implied volatility and realized measures by adopting a family of homogenous loss functions as the evaluation criteria which contain the widely used criteria as special cases. And some new and important findings related to previous studies are presented. In Section 3, a novel model average approach is proposed for combining predictors, and its forecast performance is examined. The final section concludes the paper.

2. Comparing Realized Measure and Implied Volatility

2.1. Data and forecasting models

In this section, we describe the data and recall volatilities models. There are two main type of data: daily realized measures and daily implied volatilities. The sample of observations covers the period from February 01, 2001 to September 30, 2014. This period is characterized by both calm and highly volatile periods. Then we review some volatilities models that are popular in the existing literatures and practices, and some new development in volatilities models.

2.1.1. Realized measures and models

In this study, we use the subsampled versions of realized volatilities (RV), introduced by Aït-Sahalia et al. [1], because it is among the best forecast performers (Andersen et al. [2]) and has superiorities in the accuracy of estimating asset price variation (Liu et al. [15]).

$$RV_t(m) = \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{M}(P_{t,kj/m+i} - P_{t,k(j-1)/m+i})^2} \tag{2.1}$$

where P_t, is a observation of the log price of the risky asset; $1/m$ is sampling frequency; $k = \lfloor m/M \rfloor$, rounds down to the next integer of m/M; and M is size of subsample.

In addition to RV, we use another realized measure (realized semivariance, RS), introduced by Barndorff-Nielsen et al. [3], in our analysis. These estimators are defined as

$$\begin{aligned} RS_t^+(k) &= \sqrt{\sum_{j=1}^{M} r_{t,j}^2(k) I_{[r_{t,j}(k)>0]}} \\ RS_t^-(k) &= \sqrt{\sum_{j=1}^{M} r_{t,j}^2(k) I_{[r_{t,j}(k)<0]}} \end{aligned} \tag{2.2}$$

where $r_{t,j}(k) = P_{t,kj/m} - P_{t,k(j-1)/m}$, $j = 1, \ldots, M$ is intraday returns with interval k. These estimators provide a complete decomposition of standard realized volatility.

The estimators of RV and RS were computed daily by using 5-minute intraday returns, i.e., the average interval (k/m) is 5-minute. The choice to sample prices using an approximate 5-minute window is a standard one, and is motivated by the desire to reduce the impact from microstructure noise. The data of these realized measures are provided by the database "Oxford-Man Institute's realized library" version 0.2, which has been produced by Heber et al. [12][1]. The chosen data set contains the realized volatilities for S&P 500, DJIA (Dow Jones Industria Average) and Nasdaq 100 indices. As shown in Table 1, their distributions have positive skewness and large kurtosis. This mean that the distributions are asymmetrical and leptokurtotic, especially for S&P 500 and DJIA.

Now, let us recall some models that are popular in the existing literatures and practice for using realized measures to forecast volatilities. With the advent of high-frequency data, a new pattern has occurred in which forecasting problems use simple reduced-form time series models for realized volatility measures. In this respect, the heterogeneous autoregressive (HAR) model of Corsi [7] appears to be a very successful attempt to introduce a parsimonious component approach in order to forecast daily volatilities. It can be view as parsimonious restricted versions of high-order autoregressions and given as

$$y_{h,t+h} = \alpha_0 + \alpha_d RV_t + \alpha_w RV_t^{(w)} + \alpha_m RV_t^{(m)} + \varepsilon_{t+h}, \tag{2.3}$$

where, $RV_t^{(w)} = \sum_{j=0}^{4} RV_{t-j}/5$ and $RV_t^{(m)} = \sum_{j=0}^{21} RV_{t-j}/22$ weekly and monthly realized volatility respectively, and ε_{t+h} is the random disturbance, $y_{h,t+h}$ is the h-day-ahead forecasts. Note that in our empirical analysis, we denote $y_{h,t+h}$ as RV_{t+h} for pointwise forecasting. A leverage effect can also easily be incorporated into the HAR modeling framework by including on the right-hand-side additional negative returns at different frequencies, as in Corsi and Reno [8], i.e., extended HAR model construct to $LHAR$:

$$y_{h,t+h} = \alpha_0 + \alpha_d RV_t + \alpha_w RV_t^{(w)} + \alpha_m RV_t^{(m)} + \lambda_d r_t + \lambda_t r_w^{(w)} + \lambda_m r_t^{(m)} + \varepsilon_{t+h}, \tag{2.4}$$

where $r_t^{(h)} = \min(r_t^{(h)}, 0)$, $r_t^{(w)} = \sum_{j=0}^{4} r_{t-j}/5$ and $r_t^{(m)} = \sum_{j=0}^{21} r_{t-j}/21$ denote the average weekly and monthly return respectively.

Recently, Patton and Sheppard [18] extended HAR model by decomposing the realized volatility (RV_t) into positive and negative semivariances in order to test whether signed realized variance is informative for future volatility. The extended model, denoted $HAR\text{-}RS$ henceforth, is specified as

$$y_{h,t+h} = \alpha_0 + \omega_1 RS_t^+ + \omega_2 RS_t^- + \omega_3 RV_t^{(w)} + \omega_4 RV_t^{(m)} + \varepsilon_{t+h}. \tag{2.5}$$

Within this framework, they find that disentangling the effects of negative and positive realized semivariance significantly improves forecasts of future volatility.

In our comparison, we apply recent developing models, $LHAR$ Eq. (2.4) and $HAR\text{-}RS$ Eq. (2.5), to realized measures.

[1]Further details on data preparation and data cleaning can be obtained from the documentation of the Oxford-man Institude under WWW.oxford-man.ox.ac.uk.

2.1.2. Implied volatilities

For corresponding implied volatilities, we use the implied volatility indices, VIX, VXD and VXN indices corresponding to S&P 500, DJIA and Nasdaq 100 indices respectively[2]. Since they are reported annualized, we transform them in daily ones in order to correspond to daily realized volatilities. Then we denote VIX_t, VXD_t and VXN_t as daily implied volatility. They are computed from dividing annualized implied volatility indexes by $\sqrt{252}$. We multiply the time series of realized volatilities and daily implied volatilities by 10^2 for the numerical stability of the optimization algorithms. Table 1 gives some summary statistics for these indexes. Compared with realized volatilities, their distributions are less asymmetrical and leptokurtotic, especially for Nasdaq 100.

Since we employ realized volatility that measure the volatility during trading hours as a proxy of volatility in this study and implied volatility measure the volatility for all days, the predictors based on implied volatility were given as (IV model henceforth)

$$y_{h,t+h} = c_T IV_t + \varepsilon_{t+h}, \tag{2.6}$$

where, IV_t denote implied volatility index (VIX_t, VXD_t and VXN_t), c_T is bias-adjusted factor and it can be estimated as $\hat{c}_T = \sum_{t=1}^{T-h} RV_{h,t+h} \Big/ \sum_{t=1}^{T-h} IV_t$ and T is sample size. As shown in the Table 1, the mean of realized volatilities is lower because the realized volatilities miss out on the overnight return. So employing Eq. (2.6) to adjust forecast is necessary. The IV model uses the following assumption for the multi-step-ahead forecast:

$$IV_{t+h} = IV_t \qquad \text{for } h = 1, 2, \ldots, 21. \tag{2.7}$$

As explained above, the implied volatility index is the forecast of a constant 30-day volatility implied by options. Therefore, Eq. (2.7) is consistent with the definition of the implied volatility index.

Table 1: Summary Statistics for the variables.

asset[a]	realized volatilities RV_t				Implied volatilities			
	mean	std.	skew[b]	kurt	mean	std.	skew	kurt
S&P	0.843	0.562	2.995	16.77	1.310	0.587	2.013	8.980
DJIA	0.820	0.547	3.205	19.066	1.217	0.542	1.940	8.434
Nasd	0.930	0.546	2.285	11.49	1.741	0.913	1.432	4.283

[a]S&P, DJIA and Nasd are Standard & Poor's 500 index, Dow Jones industrial average index and Nasdaq 100 index respectively.
[b]skew and kurt are skewness and kurtosis respectively.

[2]See http://www.cboe.com for more details. The VIX, VXD and VXN data are also available at the web site.

2.2. Forecast evaluation criteria

We adopt the realized volatility as the proxy for actual volatility. It is well known that realized measures are better proxies for actual volatility than squared return series. However, a better volatility proxy is still imperfect and is a noisy proxy for actual volatility. Hence, we employ the family of homogenous loss functions advocated by Patton [17] as evaluation criteria. These functions are robust to the presence of noise in realized volatility proxies. Parameterized by b, the family of loss function is defined as

$$L(v_t, rv_t, b) = \begin{cases} v_t - rv_t + rv_t \log\left(\dfrac{rv_t}{v_t}\right) & b = -1 \\ \dfrac{rv_t}{v_t} - \log\left(\dfrac{rv_t}{v_t}\right) - 1 & b = -2 \\ \dfrac{1}{b+1}\left[\dfrac{1}{b+2}\left(rv_t^{b+2} - v_t^{b+2}\right) - v_t^{b+1}(rv_t - v_t)\right] & b \notin \{-2, -1\} \end{cases} \tag{2.8}$$

where rv is a volatility measure and v is a corresponding forecast.

The above class of functions for various values of b, ranging from 2 to -2 is presented in Figure 1. This figure shows that this family of loss functions can take a wide variety of shapes, ranging from symmetric to asymmetric ($b \neq 0$), with heavier penalty either on under-prediction ($b < 0$) or over-prediction ($b > 0$). Here we use the values $b \in \{-2, -1, 0, 1, 2\}$ for forecast evaluation. This will provide more details about the forecasting performance of the predictors.

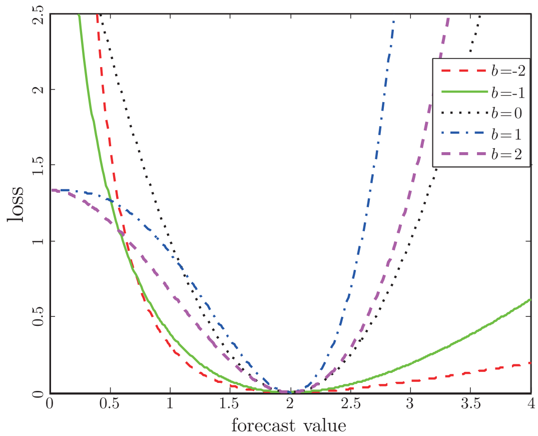

Figure 1: Loss functions for various choices of b. True value $rv = 2$ in this example, with the volatility forecast ranging between 0 and 4.

The significance of any difference in the homogenous loss function is tested via a Diebold-Mariano and West (henceforth DMW) test (see Diebold and Mariano, [9]; West [19]). A DMW statistic is computed using the difference in the losses of two models:

$$d_t = L(v_{t,1}, rv_t) - L(v_{t,2}, rv_t), \quad \text{and } DMW_T = \frac{\sqrt{T}\bar{d}_T}{\sqrt{avar(\sqrt{T}\bar{d}_T)}}, \tag{2.9}$$

where $\bar{d}_T$ is the sample mean of d_t and T is the number of forecasts. The asymptotic variance of the average is computed using a Newey-West variance estimator, with the number of lags set to 5.

2.3. Forecast evaluation

The comparison is based on in-sample and out-of-sample forecasts of daily S&P 500 index, DJIA index and Nasdaq 100 index return volatilities. For the out-of-sample forecasting, we employ the rolling window forecasting procedure with moving windows of 500 trading days (a horizon of about two years)[3]. The reason for this is that the time horizon typically used for a market risk VaR calculation and 500 trading days is a popular choice for the number of days of data used. We compare their one-step-ahead ($h = 1$) and multi-step-ahead ($h = 5$, $h = 10$ and $h = 22$) forecasting performances as well.

The forecast evaluation results of the three models are summarized in Tables 2. From the first panel, we observe that the $LHAR$ model performs the best in the in-sample fitting for all assets considered. Our results show that the predictors based on realized measures outperforms the predictors originating from implied volatility. This is in consistent with the in-sample fitting result in HP.

The second panels of Tables 2 show that the $LHAR$ model still outperforms the HAR-RS and IV models in the one-step-ahead forecast for all criteria and assets. This implies that the realized measure is more informative than implied volatility in the one-step-ahead forecast. This result is different from the results of HP, which show that the VIX index is more informative than the realized measure in the one-step-ahead forecast. The difference could be the result of the using state-of-the-art model formwork. As in their model formwork, HP employed a traditional GARCH-X model formwork. In contrast, we use recent proposed time series model formwork for volatility forecasting, which can well describe the dynamic dependencies of daily volatility.

The rest parts of Tables 2 unequivocally support the conclusion that the superiority of the implied volatility indexes is in multi-step-ahead out-of-sample forecast and become obvious with increasing forecast horizon. For the 5-step-ahead forecast, the average losses of IV model are smaller than losses of the rest models, although the DMW test shows that their forecasts are not different significantly. For the 10-step-ahead and 22-step-ahead forecasts, the IV models have the smallest average losses in almost all the cases, and the DMW test shows that the IV models are significantly better then the $LHAR$ and HAR-RS models, especially, for 22-step-ahead forecasts. Based on the discussion above, we can draw the conclusion that the decay of informational content from realized volatility is quicker than that from implied volatility. This can be at least partly explained by the different informational content of each. The information of implied volatility included the composition that reflect the option traders' judgment on the market.

Additionally, it is shown that, considering the loss function with heavier penalty on over-prediction as the evaluation criteria, the superiority of implied volatility indexes are

[3]Further details about rolling window forecasting scheme can see Han and Park [11].

not significant. In contrast, in term of loss functions with heavier penalty on under-prediction, the implied volatility indexes are superior to $LHAR$ and HAR-RS models significantly. And this become more distinct in the period of financial tsunami, as presented in Table 2A in Appendix. This is a new finding for relative studies. The new finding shows that the forecast from the time series models of realized volatility were more likely to make serious under-prediction. This can be partly explained by a very high volatility period starting in 2007. When forecasting for longer horizons, events that trigger sustained high levels of volatility, and a larger initial increase in realized volatilities only later to be followed by delayed adjustment in the series forecasts, when the larger observation becomes part of the in-sample estimation period (Martens and Zein [16]).

We also evaluate the forecasting performance of models in the period of financial tsunami starting from August 06, 2007 to January 01, 2009. For concise, we present the result in Appendix. As shown in Table 2A, the forecasting performance is similar with that for all sample period, excepting the losses become large and the superiority of implied volatility improve slightly for long forecast horizon.

3. Combination of Realized Measure and Implied Volatility

In this section, we propose a novel model average approach for combining predictors, and apply the proposed approach to combine predictors based on realized measures and implied volatility.

3.1. Methodology for model combination

Suppose for generality that there are p models (forecasts) which could be combined in order to forecast the variable of interest $y_{h,t+h}$. Denote h-step-ahead forecasts from competing models as $\{\hat{x}^{(1)}_{h,t+h}, \ldots, \hat{x}^{(p)}_{h,t+h}\}$ and consider the task of combining them in a parsimonious manner. A family of linear forecast combinations had been widely adopted starting from the seminal paper of Bates and Granger [4]. A linear forecast combination is given as

$$y_{h,t+h} = \sum_{i=1}^{p} \alpha_i \hat{x}^{(i)}_{h,t+h} + \varepsilon_{t+h}, \qquad \varepsilon_{t+h} \sim (0, \sigma^2) \tag{3.1}$$

where non-negative α_i is the weight of the i-th model with $\sum_{i=1}^{p} \alpha_i \equiv 1$. Observably, in this combination strategy the weights of candidate models are constant. Since employing more flexible weights instead of constant weights may bring more precise forecasts. A large number of forecasting literatures is devoted to seek more appropriate weighting methods, such as Bayesian model average (BMA) approach (Liu and Maheu [14]) and the empirical similarity (ES) approach (Golosnoy et al. [10]). These methods usually combine individual model forecasts based on their predictive records, i.e., the weights vary according to their forecast performance. Therefore, it is crucial to construct the weight that can response to performance of individual model effectively in combining realized measure and implied volatility.

Table 2: Forecast evaluation results.

Model	$L(-2) \times 10^2$	$L(-1) \times 10^2$	$L(0) \times 10$	$L(1) \times 10$	$L(2) \times 10$
for in-sample forecasts					
IV	4.099(7.06***)	3.684(5.74***)	0.951(4.17***)	2.313(3.70***)	8.213(3.27***)
	4.192(7.21***)	3.777(5.51***)	0.979(3.91***)	2.307(3.56***)	8.289(3.21***)
	4.037(7.73***)	4.205(7.55***)	1.139(6.61***)	0.979(5.44***)	2.053(4.43***)
HAR	2.968	2.485	0.591	1.055	3.884
	3.103	2.622	0.632	1.163	4.458
	2.820	2.582	0.640	0.546	1.220
HAR-RS	3.053(2.91**)	2.615(3.59***)	0.646(3.54***)	1.176(2.17**)	4.320(1.88*)
	3.154(1.76*)	2.723(2.80**)	0.681(3.20***)	1.259(1.99**)	4.724(1.38)
	2.902(3.28***)	2.731(4.62***)	0.699(4.62***)	0.611(4.02***)	1.371(3.36***)
for one-step-ahead forecasts ($h = 1$)					
IV	4.055(5.75***)	3.472(4.85***)	0.851(3.73***)	0.765(2.93***)	1.868(2.20**)
	4.047(5.72***)	3.425(4.59***)	0.839(3.42***)	0.771(2.64***)	1.950(2.16**)
	3.680(5.55***)	3.315(4.79***)	0.774(3.61***)	0.631(2.57**)	1.392(1.65)
$LHAR$	3.129	2.510	0.593	0.548	1.448
	3.232	2.601	0.614	0.567	1.492
	3.005	2.570	0.601	0.512	1.208
HAR-RS	3.146(0.46)	2.569(1.36)	0.631(2.00**)	0.606(1.92*)	1.615(1.38)
	3.232(0.00)	2.639(0.912)	0.649(1.81*)	0.630(2.09**)	1.720(2.03**)
	3.043(1.21)	2.654(2.03**)	0.641(2.25**)	0.561(2.01**)	1.331(1.57)
for 5-step-ahead forecasts ($h = 5$)					
IV	5.612(1.77)	4.965(1.65)	1.235(0.47)	1.107	2.657
	5.531(1.18)	4.793(0.94)	1.178(0.08)	1.062	2.602
	5.041(1.27)	4.518(0.56)	1.060	0.854	1.829
$LHAR$	5.472(1.90*)	4.821(1.98**)	1.237(1.51)	1.187(0.77)	3.100(1.09)
	5.479(1.90*)	4.732(1.73*)	1.191(0.88)	1.147(0.62)	3.067(0.90)
	4.941(1.08)	4.499(0.71)	1.088(0.78)	0.915(1.33)	2.060(1.53)
HAR-RS	5.378	4.729	1.206	1.144(0.35)	2.946(0.71)
	5.367	4.640	1.171	1.130(0.46)	3.021(0.74)
	4.884	4.455	1.076(0.45)	0.893(0.90)	1.952(0.96)
for 10-step-ahead forecasts ($h = 10$)					
IV	6.632	5.976	1.495	1.319	3.075
	6.528	5.745	1.413	1.248	2.951
	5.918	5.319	1.240	0.975	2.018
$LHAR$	6.891(1.44)	6.189(1.55)	1.550(0.75)	1.403(0.67)	3.456(0.77)
	6.805(1.55)	5.972(1.18)	1.486(0.78)	1.384(0.84)	3.577(1.02)
	6.012(0.56)	5.532(1.26)	1.315(1.51)	1.055(1.43)	2.243(1.36)
HAR-RS	6.831(1.13)	6.191(1.01)	1.565(0.83)	1.408(0.86)	3.393(0.99)
	6.722(1.10)	5.919(0.98)	1.487(1.49)	1.406(1.41)	3.692(1.16)
	5.953(0.23)	5.497(1.27)	1.311(1.82*)	1.048(1.71*)	2.193(1.50)
for 22-step-ahead forecasts ($h = 22$)					
IV	8.688	8.212	2.132	1.914	4.453
	8.502	7.817	1.982	1.756	4.069
	7.431	6.926	1.663	1.323	2.712
$LHAR$	9.531(2.51***)	9.272(2.89***)	2.519(2.40**)	2.445(1.90*)	6.324(1.67)
	9.212(2.64***)	8.817(3.25***)	2.453(2.49**)	2.581(2.11**)	7.428(1.99**)
	7.892(1.60)	7.681(2.54**)	1.885(2.61**)	1.497(2.18**)	3.022(1.71*)
HAR-RS	9.340(1.99**)	8.817(2.01**)	2.235(0.93)	1.948 (0.41)	4.4558(0.01)
9.017(1.97**)	8.451(2.60**)	2.206(2.25**)	2.080(1.86*)	5.286(1.78*)	
7.696(1.06)	7.381(2.11**)	1.773(2.31**)	1.377(1.41)	2.735(0.27)	

The loss $L(\cdot)$ is defined in Eq. (2.8) and the DMW test statistic defined in Eq. (2.9) is reported in parenthesis. Asterisks indicate rejection of the null hypothesis of equal predictability to the smallest loss model for *10%, **5% and ***1% test. The first (second and last) line of each panel is the evaluation results for S&P 500 (DJIA and Nasdaq 100).

The work of Martens and Zein [16] hint that in highly volatile periods the volatility is hard to be modeled as the models tend to be affected by structural breaks, and implied volatility may give good predictions. In contrast, when the market is running stably, stationarity of realized volatility is more reliable, and then models based on it may give desired predictions. Given this fact, we present a novel approach of combination that the weights depend on market condition.

Now let us assume that the performance of forecasts vary with the market condition and at every market condition we always can find some forecasts that behave effectively. This is in consistent with the existing empirical results that there is no dominating forecast in all case up to now. In order to describe market condition we employ a vector of variables $\mathbb{Z}_t$ characterizing the current market state. Then the weights α_i are replaced by non-negative state-based frequencies $\phi_i(\mathbb{Z}_t)$, which sum up to unity and serve as weights for the next-step forecast. The resulting forecast combination is given as

$$y_{h,t+h} = \sum_{i=1}^{p} \phi_i(\mathbb{Z}_t)\hat{x}_{h,t+h}^{(i)} + \varepsilon_{t+h}, \qquad \varepsilon_{t+h} \sim (0, \sigma^2) \tag{3.2}$$

where the $\mathbb{Z}_t$ denote current states of market condition. It usually consists of some market indexes that can reflect market condition. In term of empirical application practitioners need to select appropriate market indexes as $\mathbb{Z}$ for their implementation. It is also pivotal to configure suitable weight functions $\phi_i(\cdot)$. The proper weight functions are able to distribute the weights between forecastors effectively and achieve good combination.

According to the above-mentioned fact that in highly volatile periods the implied volatility may provide robust forecast and volatility can be modeled by realized measures well when market is stable, we choose realized volatilities $RV_{t-p-1} \cdots RV_t$ as state variables to describe market conditions, i.e., $\mathbb{Z}_t \equiv (RV_{t-p-1} \cdots RV_t)'$. Then we specify weight function $\phi(\cdot)$ for implied volatility that goes up when market become volatile and declines as market become stable. In this paper we exploit a flexible specification of the exponential weight function, which is given as

$$\phi(RV_{t-p} \cdots RV_t) = \frac{\exp(\beta(u - 1/u))}{\exp(\beta(u - 1/u)) + \exp(-\beta(u - 1/u))} \tag{3.3}$$

with

$$u = \sum_{j=0}^{p} \omega_j \frac{RV_{t-j}}{\overline{RV}} \tag{3.4}$$

that can be viewed as index of market condition and was constructed by using a natural weighting scheme where weights decline exponentially. Where $\overline{RV} = \frac{1}{T}\sum_{t=1}^{T} RV_t$ is the mean of realized volatility that reflect the average level of markets volatilities; and $\omega_j = \lambda^j(1-\lambda)/(1-\lambda^{p+1})$ is the weight used to construct market index; $(p+1)$ is the order of the model. The parameters $\lambda \in (0,1)$ deliver some information about how to describe market condition, and it can be viewed as the discount rate of historical market states. The large value of λ suggests that the model put more attention to the historical market conditions. The parameters $\beta \in \mathbb{R}$, reflect the impact from the market condition to the weight.

Obviously, the above weight function is a compound function of Eq. (3.3) and Eq. (3.4). In order to demonstrate it more clearly, we examine the function Eq. (3.3). The function Eq. (3.3) with various values of parameters β, ranging from -1.6 to 1.6 is present in Figure 2. This figure shows that this family of weight functions can take a wide variety of shapes, ranging from increasing ($\beta > 0$) to decreasing ($\beta < 0$). As shown in the Figure 2(a), when $\beta > 0$ the weight increase with the growth of variable u, and the larger the value of β is, the more quick the increasing of the weight is. From the Eq. (3.4) we can know that when the fluctuation of market goes close to the average level, the value of u approaches to 0, and the weight goes close to 0.5 accordingly. When the fluctuation of market is below (above) the average level, the value of u is smaller (larger) than 0, and the weight is smaller (larger) than 0.5 accordingly. Therefore, when market became volatile, greater (smaller) weight is given to implied volatility using the weight function with parameter "$\beta > 0$" ("$\beta < 0$").

Since the weights $\phi_i(\cdot)$ have the property $\sum_{i=1}^{p} \phi_i(\cdot) \equiv 1$, the weights of predictor based realized measures is $1 - \phi(\cdot)$. Therefore applying the new approach Eq. (3.2) to the two forecasts produces the following specification, henceforth referred as States-Based Average (SBA) model:

$$v_{h,t+h} = \theta_{1,t}\hat{v}^{(iv)}_{h,t+h} + \theta_{2,t}\hat{v}^{(rs)}_{h,t+h} + \varepsilon_{t+h}, \qquad \varepsilon_t \sim (0, \sigma^2), \tag{3.5}$$

where $\theta_{1,t} = \phi(RV_{t-p-1} \cdots RV_t)$ and $\theta_{2,t} = (1-\theta_{1,t})$; $\hat{v}^{(iv)}_{h,t+h}$ is the volatility forecasts based on implied volatility and $\hat{v}^{(rs)}_{h,t}$ is predictor from time series model of realized volatilities. And then the model parameters can be estimated by means of MLE (maximum likelihood estimation).

In our empirical analysis, we employ forecast from *HAR-RS* model for predictor using realized measures; and adopt IV model for predictor based on implied volatility, i.e., $\hat{v}^{(rs)}_{h,t+h}$ from Eq. (2.5) and $\hat{v}^{(iv)}_{h,t+h}$ from Eq. (2.6). As shown in the empirical results of Part 2.3, although the forecast from LHAR model (Eq. (2.4)) perform almost on a par with that from *HAR-RS* model (Eq. (2.5)), the *HAR-RS* model behaves more robustly. Therefore we choose *HAR-RS* model to produce predictor

3.2. Full sample estimation results

We use the same sample data of Section 2 to estimate the parameters applied in the proposed model (SBA). The SBA model given in (3.5) with order of 15 ($p = 14$), is estimated from the full sample with the maximum likelihood methodology by assuming normality for the process innovations $\{\varepsilon_{t+h}\}$[4]. The estimates of the parameters, the corresponding t-statistics in parenthesis, are reported in Table 3.

[4]For choosing the order p, after trying different orders we find that for full sample fitting and out-ofsample forecast choosing $p = 14$ can obtain better forecasting performance in most cases. Our choose based above-mentioned fact.

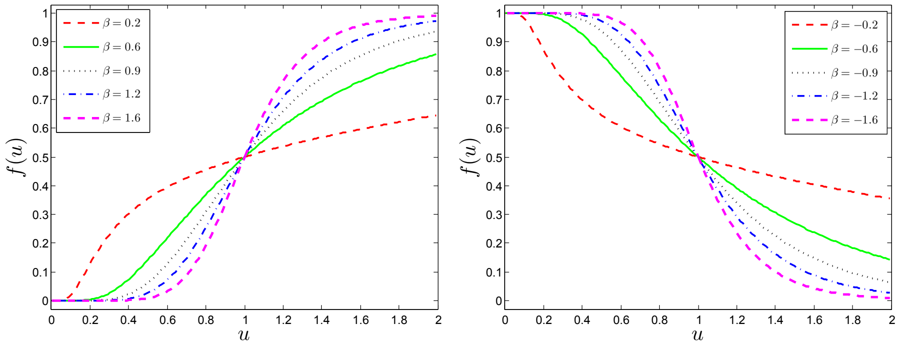

Figure 2: Weight functions for various choices of β.

The parameter estimates for the proposed model are economically reasonable. The estimates of parameters β are significantly positive for all indices. This implies that the increase of the index u enhances the weight of $\hat{v}^{(iv)}_{h,t+h}$ and consequently, raises the importance of corresponding forecast $\hat{v}^{(iv)}_{h,t+h}$ as a part of the DGP of $v_{h,t+h}$. This means that the weight function gives large weight for predictor based on implied volatility at volatile market conditions. This is economically reasonable, since in highly volatile periods the volatility is hard to be modeled as the models tend to be affected by structural breaks, and consequently, should be given small weight. In contrast, the implied volatility reflect the opinion of market participants for future volatility and can adjust with market conditions, so it is robust to provide good predictions.

Additionally, the estimate of parameter β for S&P 500 is almost equal to that for DJIA, and is distinctly smaller than that for Nasdaq 100. The Figure 3 presents the estimate of weight function for indexes considered. As shown in the figure, the shape of weight function for Nasdaq 100 is steeper than others remarkably, although they are increasing functions. This suggests that for Nasdaq 100, the forecasting performance of predictors is more sensitive to change of market conditions. This is consistent with the fact that the Nasdaq 100 consists of small firms whose prices are more affected by market condition, consequently, persistence of volatility of Nasdaq 100 is weaker than that of the other two indices. The estimate of parameters λ is greater than zero significantly. This suggests that more recent observations should be given more weight because they are more reflective of current volatilities and current macroeconomic conditions. This indicates that our weighting scheme for construction of index u is reasonable. Therefore parameter estimates for the proposed model are economically reasonable.

3.3. Evaluating combination for volatility forecast

In order to examine the effectiveness of the proposed model average approach, in this subsection we compare it with several models in both in-sample and out-of-sample forecasting performance. In particular, we compare the proposed approach with BMA and ES approach in combining implied volatility and realized measure. For this purpose we investigate the ES, BMA and SBA models. Where, the ES mode is that we apply ES

Table 3: The full sample parameter estimates and the corresponding t-statistics.

Assets	β	λ	σ
S&P 500	0.3358(2.92***)	0.6686(2.48***)	0.0659(10.63***)
DJIA	0.3526(2.98***)	0.8178(3.33***)	0.0691(8.36***)
Nasdq 100	1.3131(2.85***)	0.6682(6.63***)	0.0710(10.2***)

Asterisks indicate rejection of the null hypothesis of parameters equal to zero for *10%, **5% and ***1% test.

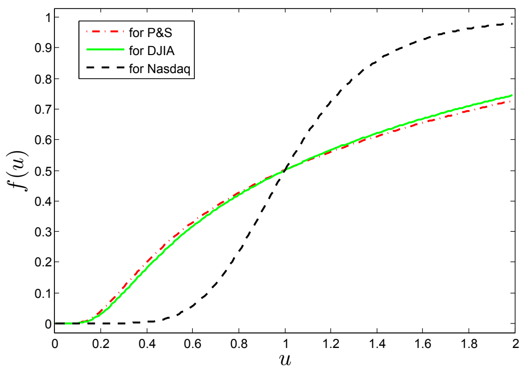

Figure 3: The estimate of weight function for each indexes.

approach of Golosnoy et al. [10] to combine $\hat{v}^{(rs)}_{h,t+h}$ and $\hat{v}^{(iv)}_{h,t+h}$ give as

$$v_{h,t+h} = \frac{\theta_{1,t}\hat{v}^{(rs)}_{h,t+h} + \theta_{2,t}\hat{v}^{(iv)}_{h,t+h}}{\theta_{1,t} + \theta_{2,t}} + \varepsilon_{t+h}, \qquad \varepsilon_t \sim (0, \sigma^2), \tag{3.6}$$

with $\theta_{1,t} = \exp\Big(-\omega_1(v_{h,t} - \hat{v}^{(rs)}_{h,t+h})^2\Big)$ and $\theta_{2,t} = \exp\Big(-\omega_2(v_{h,t} - \hat{v}^{(iv)}_{h,t+h})^2\Big)$. And the BMA model is that we use BMA methods of Liu and Maheu [14] to combine and more details can be found in their paper. Additionally, we conside the simplest approach that incorporates implied volatility into model Eq. (2.5) as an exogenous covariate (denoted by HAR-RS-IV model)

$$RV_{h,t+h} = \alpha_0 + \alpha_1 RS^+_t + \alpha_2 RS^-_t + \alpha_3 RV^{(w)}_t + \alpha_4 RV^{(m)}_t + \beta IV_t + \varepsilon_{t+h}. \tag{3.7}$$

Obviously, this model is the simplest way to combine information extracted from realized volatility and implied volatility for volatility forecasting. To run a complete forecast comparison we adopt the family of homogenous loss functions given in Eq. (2.8) again.

3.3.1. In-sample forecast evaluation

Table 4 reports the in-sample forecast evaluation result of the six models. The results show that the *HAR-RS*-IV model outperforms the rest models except ES model significantly in the within-sample fitting for all criteria considered. This implies that the implied volatility have extra information for the within-sample fitting. The finding is interesting that the simplest combination is more effective than the sophisticated models in using all information for the within-sample fitting. This finding is important for selecting model to fit volatility dynamic. But practitioners are more concerned with their out of sample forecasting performance, since many decisions are based on out-of-sample forecast. Therefore the next sections we exam their out of sample forecast extensively.

Table 4: Average values of the in-sample losses for the considered models estimated based on the full sample information.

Model	$L(-2) \times 10^2$	$L(-1) \times 10^2$	$L(0) \times 10$	$L(1) \times 10$	$L(2) \times 10$
Estimation results for S&P 500					
IV	4.101(8.98***)	3.682(6.74***)	0.950(4.83***)	0.876(3.69***)	2.117(2.75***)
HAR-RS	3.054(2.56***)	2.613(3.96***)	0.646(3.97***)	0.587(3.16***)	1.449(2.48***)
HAR-RS-IV	2.970(0.27)	2.514	0.616	0.558	1.374
ES	2.960	2.524(0.30)	0.622(0.52)	0.566(0.67)	1.403(0.97)
BMA	3.055(2.59***)	2.614(3.96***)	0.645(3.96***)	0.586(3.16***)	1.448(2.49***)
SBA	3.164(3.44***)	2.949(3.56***)	0.810(3.39***)	0.792(3.06***)	1.992(2.49***)
Estimation results for DJIA					
IV	3.780(6.76***)	3.315(5.73***)	0.774(4.16***)	0.631(2.76***)	1.392(1.52)
HAR-RS	3.043(2.13***)	2.654(3.07***)	0.641(3.00***)	0.561(2.27***)	1.331(1.44)
HAR-RS-IV	2.950	2.522	0.598	0.521	1.253
ES	2.979(0.72)	2.600(1.70)	0.630(1.66)	0.554(1.32)	1.315(0.89)
BMA	3.075(2.87***)	2.681(3.66***)	0.647(3.40***)	0.569(2.65***)	1.357(1.93*)
SBA	3.358(4.37***)	2.984(3.88***)	0.715(3.04***)	0.600(2.15**)	1.352(1.13)
Estimation results for Nasdaq100					
IV	4.045(9.53***)	4.199(8.77***)	1.132(7.64***)	0.969(6.29***)	2.026(4.57***)
HAR-RS	2.909(4.04***)	2.728(3.94***)	0.694(2.31**)	0.603(1.41)	1.349(1.10)
HAR-RS-IV	2.836(0.66)	2.668	0.682	0.595	1.334
ES	2.825	2.670(0.09)	0.685(0.58)	0.598(0.65)	1.342(0.68)
BMA	2.907(3.95***)	2.724(3.75***)	0.693(2.12**)	0.602(1.27)	1.348(1.00)
SBA	3.512(6.93***)	3.687(6.76***)	1.018(6.15***)	0.896(5.26***)	1.923(3.97**)

Asterisks indicate rejection of the null hypothesis of equal predictability to the smallest loss model for *10%, **5% and ***1% test.

3.3.2. Out-of-sample forecast evaluation

For the out-of-sample forecasting, we employ the rolling window forecasting procedure with moving windows of 500 trading days. This is same to Section 2.3, where we give the reason for the choice. We conduct one-step-ahead ($h = 1$) as well as multi-step-ahead ($h = 5$, $h = 10$ and $h = 22$) forecasts. Note that for SBA and ES models there exist some difference in forecasting methodology from other models. Since they need to use the forecasts of component models, for each rolling window the forecasting procedure must be divided into two steps. Firstly, we obtain 250 out-of-sample forecasts for each component model, employing the rolling window forecasting procedure with moving sub-windows of 250 trading days confined in the moving window of 500 trading days[5]. Secondly, using out-of-sample forecasts obtained in the first step, we estimate the parameters of the SBA and ES models, and then use the estimation values to obtain out-of-sample forecasts. For the order of the SBA model we adopt $p = 14$ to conduct out-of-sample forecast.

Table 5: Average losses of one-step-ahead forecasts.

Model	$L(-2) \times 10^2$	$L(-1) \times 10^2$	$L(0) \times 10$	$L(1) \times 10$	$L(2) \times 10$
Estimation results for S&P 500					
IV	4.055(7.61***)	3.472(6.12***)	0.852(4.24***)	0.766(2.86***)	1.868(1.94*)
HAR-RS	3.146(3.72***)	2.569(3.64***)	0.631(2.62***)	0.606(1.97*)	1.615(1.73)
HAR-RS-IV	3.006	2.452	0.599	0.569	1.508
ES	3.112(2.55***)	2.528(2.19***)	0.617(1.36)	0.589(1.00)	1.566(0.89)
BMA	3.141(3.58***)	2.567(3.51***)	0.630(2.58***)	0.605(1.96*)	1.613(1.73)
SBA	3.592(5.39***)	3.092(4.27***)	0.778(3.17***)	0.722(2.32**)	1.806(1.65)
Estimation results for DJIA					
IV	4.048(7.52***)	3.425(5.89***)	0.839(3.96***)	0.771(2.68***)	1.950(1.86*)
HAR-RS	3.232(2.96***)	2.639(2.79***)	0.649(2.43**)	0.630(2.27**)	1.720(2.11**)
HAR-RS-IV	3.119	2.538	0.616	0.584	1.559
ES	3.196(1.95*)	2.596(1.74)	0.634(1.50)	0.615(1.58)	1.680(1.72)
BMA	3.224(2.79***)	2.636(2.68***)	0.649(2.40**)	0.630(2.27**)	1.720(2.12**)
SBA	3.581(4.80***)	3.045(3.82***)	0.764(2.85***)	0.724(2.14**)	1.874(1.60)
Estimation results for Nasdaq100					
IV	3.780(6.76***)	3.315(5.73***)	0.774(4.16***)	0.631(2.76***)	1.392(1.52)
HAR-RS	3.043(2.13**)	2.654(3.07***)	0.641(3.00***)	0.561(2.27**)	1.331(1.44)
HAR-RS-IV	2.950	2.522	0.598	0.520	1.253
ES	2.979(0.72)	2.600(1.70)	0.630(1.66)	0.554(1.32)	1.315(0.89)
BMA	3.075(2.87***)	2.681(3.66***)	0.647(3.40***)	0.569(2.65***)	1.357(1.93*)
SBA	3.358(4.37***)	2.984(3.88***)	0.715(3.04***)	0.600(2.15**)	1.352(1.13)

Asterisks indicate rejection of the null hypothesis of equal predictability to the smallest loss model for *10%, **5% and ***1% test.

Employing robust loss function given in Eq. (2.8), the out-of-sample forecast evaluation results of the interesting models are summarized in Tables 5-8. From the Table 5, we observe that the results concerning one-step-ahead forecast performance is similar to that for within-sample fitting. The *HAR-RS*-IV model maintains preeminence in one-step-ahead forecast performance, although its superiority have weakened slightly according to corresponding *DMW* tests. This shows that it is the most effective model in using all information for the one-step-ahead forecast. This is reasonable because the model with good fitting usually can bring favorable out-of-sample forecast when the forecasting horizon is narrow enough.

As shown in the Table 6, for 5-step-ahead forecast the average loss of the proposed model (SBA) is the smallest in most cases. This suggests that SBA model has taken

[5]The size of subwindows is not always equal to the number of the out-of-sample forecasts within each rolling window. This can be adjusted according to the number of parameters that need to be estimated.

over merits from HAR-RS-IV model (which performs the best in within-sample fitting and one-step-ahead forecasting) and outperforms the rest of the models in the 5-step-ahead forecast evaluation. This implies that the SBA model is more effective than the rest models in utilizing information from implied volatility and realized measures for the 5-step-ahead forecast.

The Table 7 report the evaluation results of 10-step-ahead forecast. As reported in the table, the average loss of the SBA model maintain the smallest for all criteria and assets, and the t-value of DMW test for null hypothesis of equal predictability to the smallest loss model have reached significant level, especially for that based on the loss $L(-2)$, $L(-1)$ and $L(0)$. This suggests that the 10-step-ahead forecast of the SBA model is significantly better than that of the rest models. Comparing Table 7 with Table 6, we observe that for the 10-step-ahead forecast the superiority of SBA model are more obvious. This implies that in utilizing information from implied volatility and realized measures for volatility forecast, the SBA model becomes more effective as the forecast horizon increase to 10 trading days. As we have mentioned in the introduction, 10-trading-day-ahead volatility forecast is key important for risk management. Therefore these improvement will benefit the field of risk management.

The Table 8 present the evaluation results of 22-step-ahead forecast. The results are similar to that of 10-step-ahead forecast. The proposed model retains the lead in the 22-step-ahead forecast and was followed by ES model closely. And the DMW test shows that their forecasts are not significantly different. This is not surprising, because the IV model dominates HAR-RS model for 22-step-ahead forecast (see the empirical results of Section 2.3), and both SBA and ES models give larger weight to IV model than HAR-RS model. So they perform similar for 22-step-ahead forecast.

Additionally, considering the average losses of $L(1)$, $L(2)$ from Table 6$-$8 and according to the asymmetric penalty feature of the loss functions, the IV model was able to reduce overestimation for the multi-step-ahead forecast. This is a new finding for relative empirical study. The finding is pivotal for improving volatility forecasts.

In Sum, firstly we find that the HAR-RS-IV model outperforms the rest models significantly for the within-sample fitting and the one-step-ahead forecast. The finding is interesting for the fact that the simplest combination is more effective than the sophisticated models in using all information for the within-sample fitting and the one-step-ahead forecast.

Secondly, there is a new finding for relative empirical study that IV model is able to reduce overestimation for the multi-step-ahead forecast. The finding is pivotal for improving volatility forecasts. Finally, the proposed model performs the best for multi-step-ahead out-of-sample forecast, especially for 10-step- ahead and 22-step-ahead forecast. This implies that the proposed model is able to utilize information from implied volatility and realized measures effectively for multi-step-ahead forecast. Since 10-trading-day-ahead volatility forecast is extremely important for risk management, these improvement will benefit the field of risk management remarkably.

Table 6: Average losses of 5-step-ahead forecasts.

Model	$L(-2) \times 10^2$	$L(-1) \times 10^2$	$L(0) \times 10$	$L(1) \times 10$	$L(2) \times 10$
Estimation results for S&P 500					
IV	5.612(3.65***)	4.965(2.57***)	1.235(6.82***)	1.107(5.02***)	2.657(3.03***)
HAR-RS	5.378(3.84***)	4.729(2.45**)	1.206(0.35)	1.144(0.69)	2.946(0.84)
HAR-RS-IV	5.232	4.638	1.203(0.27)	1.167(0.79)	3.075(0.98)
ES	5.342(1.83*)	4.745(1.13)	1.202(0.80)	1.099(1.06)	2.675(1.14)
BMA	5.410(4.32***)	4.773(3.12***)	1.217(0.54)	1.149(0.74)	2.950(0.85)
SBA	5.367(1.75)	4.737(0.91)	1.186	1.075	2.612
Estimation results for DJIA					
IV	5.531(2.87***)	4.793(1.70)	1.178(6.04***)	1.062(4.02***)	2.602(2.21**)
HAR-RS	5.369(3.18***)	4.640(2.44**)	1.171(0.57)	1.130(0.74)	3.021(0.84)
HAR-RS-IV	5.233	4.554	1.164(0.47)	1.139(0.79)	3.071(0.92)
ES	5.311(1.47)	4.668(1.56)	1.204(1.36)	1.175(1.40)	3.127(1.34)
BMA	5.376(3.56***)	4.661(2.96***)	1.177(0.65)	1.133(0.76)	3.023(0.84)
SBA	5.277(0.58)	4.565(0.10)	1.129	1.030	2.556
Estimation results for Nasdaq100					
IV	5.041(7.07***)	4.518(6.80***)	1.060(5.64***)	0.854(3.70***)	1.829(1.97*)
HAR-RS	4.884(1.56)	4.455(1.77)	1.076(1.64)	0.893(1.36)	1.952(1.14)
HAR-RS-IV	4.777(0.30)	4.358(0.60)	1.070(1.00)	0.925(1.22)	2.148(1.34)
ES	4.772(0.45)	4.349(1.22)	1.043(1.43)	0.855(1.33)	1.847(1.19)
BMA	4.933(2.40**)	4.529(2.40**)	1.116(1.90*)	0.965(1.60)	2.229(1.42)
SBA	4.759	4.309	1.026	0.838	1.812

Asterisks indicate rejection of the null hypothesis of equal predictability to the smallest loss model for *10%, **5% and ***1% test.

Table 7: Average losses of 10-step-ahead forecasts.

Model	$L(-2) \times 10^2$	$L(-1) \times 10^2$	$L(0) \times 10$	$L(1) \times 10$	$L(2) \times 10$
Estimation results for S&P 500					
IV	6.632(3.45***)	5.976(3.61***)	1.495(3.22***)	1.319(2.17**)	3.075(1.18)
HAR-RS	6.831(3.52***)	6.143(2.97***)	1.544(1.88*)	1.400(1.38)	3.424(1.20)
HAR-RS-IV	6.619(1.93*)	6.045(2.08**)	1.560(1.80*)	1.458(1.57)	3.671(1.32)
ES	6.543(3.23***)	5.915(3.39***)	1.482(2.55***)	1.310(1.49)	3.057(0.68)
BMA	6.740(3.34***)	6.098(2.87***)	1.546(1.95*)	1.410(1.51)	3.458(1.29)
SBA	6.446	5.809	1.456	1.292	3.033
Estimation results for DJIA					
IV	6.528(4.05***)	5.745(4.28***)	1.413(4.01***)	1.248(2.93***)	2.951(1.69)
HAR-RS	6.722(3.41***)	5.919(2.60***)	1.487(1.55)	1.406(1.35)	3.692(1.25)
HAR-RS-IV	6.539(2.43**)	5.829(2.12**)	1.490(1.58)	1.426(1.46)	3.761(1.31)
ES	6.463(3.67***)	5.710(4.27***)	1.406(3.71***)	1.243(2.40**)	2.941(1.27)
BMA	6.627(3.45***)	5.886(2.58***)	1.492(1.62)	1.416(1.42)	3.715(1.29)
SBA	6.330	5.577	1.375	1.222	2.912
Estimation results for Nasdaq100					
IV	5.918(6.03***)	5.319(5.23***)	1.240(3.88***)	0.975(2.36**)	2.018(1.36)
HAR-RS	5.953(3.24***)	5.497(3.52***)	1.311(2.88***)	1.048(2.08**)	2.194(1.66)
HAR-RS-IV	5.848(2.80***)	5.391(2.59***)	1.308(1.91**)	1.094(1.64)	2.436(1.40)
ES	5.695(2.98***)	5.196(2.95***)	1.225(2.24**)	0.970(1.31)	2.011(0.66)
BMA	5.872(2.74***)	5.430(2.82***)	1.321(2.06**)	1.112(1.63)	2.504(1.34)
SBA	5.630	5.138	1.214	0.965	2.007

Asterisks indicate rejection of the null hypothesis of equal predictability to the smallest loss model for *10%, **5% and ***1% test.

Table 8: Average losses of 5-step-ahead forecasts.

Model	$L(-2)\times 10^2$	$L(-1)\times 10^2$	$L(0)\times 10$	$L(1)\times 10$	$L(2)\times 10$
Estimation results for S&P 500					
IV	8.688(2.41**)	8.212(2.09**)	2.132(2.49**)	1.914(3.15***)	4.453(2.98***)
HAR-RS	9.339(3.94***)	8.817(3.38***)	2.235(2.09**)	1.948(0.87)	4.456(0.33)
HAR-RS-IV	9.418(3.42***)	8.976(3.12***)	2.285(2.33***)	1.986(1.27)	4.518(0.58)
ES	8.517(0.56)	8.123(0.70)	2.120(1.00)	1.912(1.16)	4.468(1.19)
BMA	9.054(3.12***)	8.591(2.83***)	2.196(1.66)	1.930(0.64)	4.439(0.25)
SBA	8.486	8.088	2.103	1.885	4.384
Estimation results for DJIA					
IV	8.502(2.81***)	7.817(2.41**)	1.982(1.99**)	1.756(1.20)	4.069(0.26)
HAR-RS	9.017(3.91***)	8.451(4.07***)	2.206(2.79***)	2.080(2.07**)	5.286(1.71)
HAR-RS-IV	9.150(3.99***)	8.641(4.12***)	2.260(3.06***)	2.124(2.19**)	5.390(1.76)
ES	8.235	7.677(0.32)	1.965(0.89)	1.750(0.75)	4.064(0.18)
BMA	8.842(3.29***)	8.344(3.62***)	2.194(2.66***)	2.078(2.05**)	5.289(1.71)
SBA	8.242(0.13)	7.663	1.955	1.740	4.058
Estimation results for Nasdaq100					
IV	7.431(4.85***)	6.926(3.67***)	1.663(2.56***)	1.324(2.07**)	2.712(2.08**)
HAR-RS	7.696(3.89***)	7.381(4.28***)	1.773(3.56***)	1.377(1.85***)	2.735(0.55)
HAR-RS-IV	7.642(3.52***)	7.306(3.79***)	1.760(3.26***)	1.378(1.98**)	2.765(1.00)
ES	7.0883(0.19)	6.7648(0.65)	1.6507(1.30)	1.323(2.08**)	2.713(2.45**)
BMA	7.511(3.28***)	7.204(3.64***)	1.737(2.83***)	1.358(1.33)	2.713(0.27)
SBA	7.080	6.741	1.641	1.314	2.692

Asterisks indicate rejection of the null hypothesis of equal predictability to the smallest loss model for $^{*}10\%$, $^{**}5\%$ and $^{***}1\%$ test.

4. Conclusion

In this paper we reinvestigate the forecasts of volatility obtained from the implied volatility and the series models of realized volatilities. The differences between our work and previous studies are that we adopt a family of homogenous loss functions as the evaluation criteria and use wider sample data. To run a fair and complete forecast comparison we employ a family of homogenous loss functions as the evaluation criteria. The comparison is based on in-sample and out-of-sample forecasts of daily S&P 500 index, DJIA index and Nasdaq 100 index return volatilities for recent long period.

Our empirical results of comparison show that, predictors based on realized measures are superior to that derived from implied volatility for within-sample fitting. This is consistent with the previous studies. For out-of-sample forecast, the realized measures are more informative than implied volatility in the one-step-ahead forecast. Our result is different from the results of HP. The reason for the difference may be due to the fact that we employ state of the art model of realized measures. Whereas the superiority of the implied volatility indexes is in multi-step-ahead forecast and becomes obvious with increasing forecast horizon. Additionally, it is shown that, considering the loss function with heavier penalty on over-prediction as the evaluation criteria, the implied volatility indexes are not able to outperform LHAR and *HAR-RS* models significantly. In contrast, in term of loss functions with heavier penalty on under-prediction, the implied volatility indexes are superior to LHAR and *HAR-RS* models, and become more distinct in the period of financial tsunami. These are new findings that add to the findings of previous studies. And it is also of importance for risk management.

Secondly, in view of above empirical result that two types of predictors show complementary strengths in the volatility forecasting performance, we propose a new model average approach to combine them in order to obtain more excellent forecast in spirit of complementary advantages. The new approach sets the weights of the com- ponent models depending on the market conditions. Consequently, we use realized volatility to construct an index describing market condition, and specify a flexible weight function. And the obtained estimation results confirm that the proposed approach is economically reasonable.

Finally, adopting a family of homogenous loss functions as the evaluation criteria, we compare proposed approach with naive model (HAR-RS-IV), the empirical similarity (ES) approach and Bayesian models average approach in combining realized measures and implied volatility for volatility prediction. There are some interesting findings in our empirical results: Firstly we find that the HAR-RS-IV model outperforms the rest models significantly for the within-sample fitting and the one-step-ahead forecast. The finding is interesting for the fact that the simplest combination is more effective than the sophisticated models in using all information for the within-sample fitting and the one-step-ahead forecast. Secondly, there is a new finding for relative empirical study that IV model was able to reduce overestimation for the multi-step-ahead forecast. The finding is pivotal for improving volatility forecasts. Finally, the proposed model performs the best for multi-step-ahead out-of-sample forecast, especially for 10-step-ahead and 22-step-ahead forecast. This implies that the proposed model is able to utilize information from implied volatility and realized measures effectively for multi-step-ahead forecast. Since 10-trading-day-ahead volatility forecast is important for risk management, these improvement will benefit the field of risk management remarkably, especially for forecasting the VaR of market risk. Obviously, our empirical results are consistent with the existing empirical result that there is no dominating forecast up to now.

Acknowledgements

The authors appreciate the financial support of the Natural Science Foundation of Ningbo project, No.2018A610130; National statistical science research program, 2019 LY71; Ningbo Soft Science Foundation No. 2017A10113; Fujian Education Department, No. JT180444High-Level Personnel Fund of Xiamen University of Technology, No. YKJ15031R. The authors thank the anonymous referees for careful reading and suggestions.

Appendix

Table 2A: Forecast evaluation results for period of financial tsunami.

Model	$L(-2) \times 10^2$	$L(-1) \times 10^2$	$L(0) \times 10$	$L(1) \times 10$	$L(2) \times 10$
for in-sample forecasts					
IV	6.197(3.38***)	9.827(3.22***)	3.965(2.87**)	14.59(2.89**)	58.31(2.62**)
	5.985(3.40***)	9.333(3.04***)	3.783(2.65**)	13.92(2.79**)	55.69(2.56**)
	7.433(4.17***)	10.58(3.88***)	3.700(3.27***)	11.63(3.30***)	39.18(2.83**)
LHAR	3.300	4.775	1.870	5.615	25.43
	3.300	4.679	1.811	5.759	26.76
	3.446	4.872	1.793	4.411	17.75
HAR-RS	3.452(1.47)	5.247(2.27**)	2.165(2.47**)	6.163(1.25)	27.82(1.18)
	3.390(0.84)	5.080(1.93*)	2.092(2.35**)	6.032(0.74)	27.50(0.44)
	3.727(2.83**)	5.468(3.09***)	2.071(2.82**)	5.082(1.69*)	20.22(1.52)
for one-step-ahead forecasts ($h = 1$)					
IV	5.358(3.24***)	8.145(3.20***)	3.172(2.88**)	3.871(2.36**)	11.26(1.68*)
	5.288(3.29***)	7.855(3.12***)	3.071(2.78**)	3.807(2.46**)	11.23(2.14**)
	4.919(3.14***)	7.009(2.99***)	2.493(2.29**)	2.753(1.50)	7.392(1.65)
LHAR	3.402	4.886	1.951	2.608	8.521
	3.498	4.920	1.908	2.463	7.760
	3.485	4.985	1.875	2.246	6.399
HAR-RS	3.590(1.37)	5.468(2.32**)	2.285(2.36**)	3.085(1.84*)	9.850(1.21)
	3.566(0.47)	5.353(1.82*)	2.222(2.37**)	2.998(2.3**)	9.612(2.06**)
	3.777(2.36**)	5.639(2.61**)	2.179(2.19**)	2.615(1.69*)	7.310(1.23)
for 5-step-ahead forecasts ($h = 5$)					
IV	8.129(0.01)	12.49(0.21)	4.876	5.906	16.82
	7.988(0.03)	11.94(0.23)	4.621	5.593	15.92
	7.268	10.47	3.715	4.025	10.22
LHAR	8.437(1.30)	12.69(1.42)	5.104(0.49)	6.596(0.82)	20.31(1.04)
	8.311(1.40)	11.96(0.95)	4.761(0.23)	6.220(0.57)	19.45(0.85)
	7.830(1.10)	10.99(0.87)	3.938(0.86)	4.438(1.03)	11.89(1.22)
HAR-RS	8.125	12.31	4.924	6.286(0.46)	19.10(0.71)
	7.968	11.70	4.686	6.118(0.46)	19.09(0.71)
	7.710(0.94)	11.06(1.02)	3.953(0.99)	4.325(0.90)	11.07(0.82)
for 10-step-ahead forecasts ($h = 10$)					
IV	10.47	16.02	6.131	7.175	19.61
	10.34	15.40	5.811	6.743	18.31
	9.104	13.10	4.546	4.730	11.46
LHAR	11.89(1.94*)	17.43(1.37)	6.572(0.86)	7.891(0.77)	22.84(0.87)
	11.36(1.26)	16.26(0.72)	6.265(0.61)	7.822(0.89)	23.48(1.13)
	10.08(1.23)	13.86(0.79)	4.786(0.70)	5.116(0.91)	12.94(1.15)
HAR-RS	12.29(2.34**)	17.83(1.97**)	6.718(1.38)	8.007(1.17)	22.82(1.88*)
	11.45(1.39)	16.56(1.08)	6.447(1.01)	8.129(1.18)	24.64(1.33)
	10.26(1.95*)	15.41(1.87*)	4.986(1.64)	5.232(1.47)	12.81(1.38)
for 22-step-ahead forecasts ($h = 22$)					
IV	16.59	25.69	9.76	11.20	29.84
	16.02	24.22	9.03	10.17	26.59
	13.53	19.91	6.92	7.06	16.58(0.12)
LHAR	21.72(3.67***)	33.03(3.25***)	12.87(2.60***)	15.73(2.12**)	45.83(1.85*)
	19.84(3.32***)	30.40(2.99***)	12.61(2.34**)	16.93(2.10**)	54.49(2.00**)
	17.27(2.62***)	24.22(2.33**)	8.141(2.01**)	8.127(1.73**)	18.78(1.73*)
HAR-RS	20.85(3.04***)	29.65(2.42***)	10.55(1.51)	11.61(0.66)	30.35(0.26)
	719.29(3.78***)	28.10(2.64***)	10.67(2.14**)	12.83(1.87*)	36.80(1.81*)
	16.27(2.67***)	22.36(2.29**)	7.354(1.52)	7.215(0.52)	16.49

The loss $L(\cdot)$ is defined in (2.8) and the DMW test statistic defined in (2.9) is reported in parenthesis. Asterisks indicate rejection of the null hypothesis of equal predictability to the smallest loss model for *10%, **5% and ***1% test. The first (second and last) line of each panel is the evaluation results for S&P 500 (DJIA and Nasdaq 100).

References

[1] Aït, S., Mykland, P. A. and Zhang, L. (2005). *How often to sample a continuous time process in the presence of market microstructure noise*, Review of Financial Studies, Vol.18, 351-416.

[2] Andersen, T. G., Bollerslev, T. and Meddahi, N. (2011). *Realized volatility forecasting and market microstructure noise*, Journal of Econometrics, Vol.160, 220-234.

[3] Bollerslev, T., Russell, J. and Watson, M. (2010). Volatility and Time Series Econometrics, Oxford University Press.

[4] Bates, J. M. and Granger, C. W. J. (1969). *The combination of forecasts*, Operation Research Quarterly, Vol.20, 451-468.

[5] Blair, B. J., Poon, S. H and Taylor, S. J. (2001). *Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns*, Journal of Econometrics, Vol.105, 5-26.

[6] Clemen, R. T. (1989). *Combining forecasts: a review and annotated bibliography*, International Journal of Forecasting, Vol.5, No.4, 559-583.

[7] Corsi, F. (2009). *A simple approximate long-memory model of realized volatility*, Journal of Financial Econometrics, Vol.7, No.2, 174-196.

[8] Corsi, F. and Renò, R. (2012). *Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling*, Journal of Business & Economic Statistics, Vol.30, No.3, 368-380.

[9] Diebold, F. X. and Mariano, R., S. (1995). *Comparing predictive accuracy*, Journal of Business and Economic Statistics, Vol.13, 253-263.

[10] Golosnoy, V., Hamid, A. and Okhrin, Y. (2014). *The empirical similarity approach for volatility prediction*, Journal of Banking and Finance, Vol.40, 321-329.

[11] Han, H. and Park, M. D. (2013). *Comparison of realized measure and implied volatility in forecasting volatility*, Journal of Forecasting, Vol.32, No.6, 522-533.

[12] Heber, G., Lunde, A., Shephard, N. and Sheppard, K. K. (2014). OMIs Realised Library, Version 0.2. Oxford-Man Institute: University of Oxford.

[13] Koopman, S. J., Jungbacker, B. and Hol, E. (2005). *Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements*, Journal of Empirical Finance, Vol.12, 445-475.

[14] Liu, C. and Maheu, J. M. (2009). *Forecasting realized volatility: a Bayesian model-averaging approach*, Journal of Applied Econometrics, Vol.24, No.5, 709-733.

[15] Liu, L. Y., Patton, A. J. and Sheppard, K. (2015). *Does anything beat* 5- *minute RV? A comparison of realized measures across multiple asset classes*, Journal of Econometrics, Vol.187, 293-311.

[16] Martens, M. and Zein, J. (2004). *Predicting financial volatility: High-frequency time-series forecasts vis-à-vis implied volatility*, Journal of Futures Markets, Vol.24, No.11, 1005-1028.

[17] Patton, A. J. (2011). *Volatility forecast comparison using imperfect volatility proxies*, Journal of Econometrics, Vol.160, 246-256.

[18] Patton, A. J. and Sheppard, K. (2015). *Good volatility, bad volatility: Signed jumps and the persistence of volatility*, Review of Economics and Statistics, Vol.97, No.3, 683-697.

[19] West, K. D. (2006). *Chapter* 3 *forecast evaluation*, Handbook of Economic Forecasting, Vol.1, No.05, 99-134.

School of Science, Ningbo University of Technology, Ningbo, Zhejiang Province, 315211, P.R. China.

E-mail: shiyafonglf@126.com

Major area(s): Financial time series, Risk management.

Zhejiang Pharmaceutical College, Ningbo, Zhejiang, 315100, P.R. China.

E-mail: shiyan-long@hotmail.com

Major area(s): Applied mathematics.

Guangdong Guangya High School, Gangzhou, Guangdong, 510160, P.R. China.

E-mail: 18002293927@189.com

Major area(s): Numerical mathematics.

School of Applied Mathematics, Xiamen University of Technology, Xiamen, Fujian, 361024, P.R. China.
E-mail: zhunenghui163@163.com
Major area(s): Applied statistics.

Faculty of Business, University of Nottingham Ningbo China, Ningbo, Zhejiang, 315100, P.R. China.
E-mail: yingtingting1028@hotmail.com (*Corresponding author)
Major area(s): Corporate finance.

Faculty of Economics, Shanxi University of Finance and Economics, Taiyuan, Shanxi, 030006, P.R. China.
E-mail: juyan255@126.com
Major area(s): Game, match.

(Received July 2018; accepted March 2019)