



Forecasting Realized Volatility Dynamically Based on Adjusted Dynamic Model Averaging (AMDA) Approach: Evidence from China's Stock Market

Ping Yuan

Antai College of Economics & Management, Shanghai Jiaotong University, Shanghai, China

Email address:

alexpiggy1314@163.com

To cite this article:

Ping Yuan. Forecasting Realized Volatility Dynamically Based on Adjusted Dynamic Model Averaging (AMDA) Approach: Evidence from China's Stock Market. *International Journal of Accounting, Finance and Risk Management*. Vol. 4, No. 2, 2019, pp. 44-53.

doi: 10.11648/j.ijafirm.20190402.11

Received: April 25, 2019; **Accepted:** June 2, 2019; **Published:** June 20, 2019

Abstract: In this study, we forecast the realized volatility of the CSI 300 index using the heterogeneous autoregressive model for realized volatility (HAR-RV) and its various extensions. Our models take into account the time-varying property of the models' parameters and the volatility of realized volatility. The adjusted dynamic model averaging (ADMA) approach, is used to combine the forecasts of the individual models. Different from DMA method, the least and second least probability of particular models are excluded from the process of averaging the forecasts across the different models when using ADMA method. Our empirical results suggest that ADMA can generate more accurate forecasts than DMA method and alternative strategies. Models that use time-varying parameters have greater forecasting accuracy than models that use the constant coefficients. Time-varying parameter (TVP) models can generate more accurate forecasts than constant coefficient models in China's stock market. The robustness test also indicates that the prediction accuracy of these DMA and ADMA models based on different parameters is higher than that of most single models, which further proves the effectiveness of the multi-model realized volatility prediction model based on dynamic averaging method in the prediction effect. These findings indicate the importance of considering parameter change and model specification change in volatility forecasting.

Keywords: CSI 300 Index, Realized Volatility, Adjusted Dynamic Model Averaging, Time-varying Parameters

1. Introduction

It is widely acknowledged that volatility forecasting plays a crucial role in asset pricing, portfolio optimization, and risk management [1], and many studies have demonstrated that realized volatility (RV)-based models, fully utilizing high frequency data, are more effective than traditional GARCH models. The introduction of realized volatility (RV) or realized variance opens a new era of forecasting and modeling in financial market volatility [2-5]. This nonparametric measure of volatility, defined as the sum of all available intraday squared returns, is considered a better proxy of unobserved actual volatility than squared daily returns. Therefore, it is not surprising that wide-range volatility models are developed to capture and forecast the dynamics of RV [6-12, 13-15].

Among the models for RV, heterogeneous autoregressive

realized volatility (HAR-RV) developed by Corsi [13] is one of the most popular. HAR-RV is a predictive regression which takes lagged daily, weekly and monthly RV as the explanatory variable for future RV. The simulation and empirical results in Corsi [13] suggest that the simple HAR-RV can accommodate some 'stylized facts' in financial volatility such as long memory, multiscaling and fat-tail distribution. Given the good performance of HAR-RV in modeling and forecasting volatility, a series of meaningful extensions have been proposed based on different decomposition methodologies of RV. For example, Andersen et al. [6] decomposes daily RV into a continuous sample path, and a significant jump component and their proposed HAR-RV-CJ model uses these two components as the predictors of future RV. Corsi et al. [14] improves the decomposition method of Andersen et al. [6] by introducing a new threshold of bipower variation for estimating the jump component and develops a HAR-RV-TCJ

model accordingly. So an important improvement on the RV-based models is by adding the jump component into the models, lots of articles put their attention on defying jump process and discovering its impact [16-19].

In this study, we forecast the realized volatility of the CSI 300 index using the HAR-RV model and its extensions. We contribute to the literature in following dimensions. First, we consider time-variation in parameters of volatility models. Due to many factors such as business cycles, extreme events, and economic policies, the statistical property of volatility (e.g., volatility persistence) undergoes frequent structural breaks or switches between different regimes [20, 21]. Because of these factors, the statistical property of volatility is likely to change over time. To investigate the effects of structural breaks on the predictive ability, we consider two types of volatility models depending on the differences in parameters. The first is the constant coefficient (CC) models that assume no structural break in volatility dynamics. The second is the time-varying parameter (TVP) models implying that the structural breaks occur at each point of time. Since there are some empirical results show that the time-varying parameter models have advantages on the volatility forecasting, our article aims to prove its effectiveness on the Chinese stock market.

Second, we use an adjusted dynamic model averaging (ADMA) approach to combine the forecasts of individual models. The motivation for this approach is the argument that the predictability of a single model is very unstable over time [22]. Choosing the forecasts of an individual model ignores this model uncertainty, understates the forecasting risks, and is likely to result in poor predictions [23, 24]. For example, to capture the effects of jump components on realized volatility, Andersen et al. [6] add a jump component to the HAR-RV model, making it the HAR-RV-J model. We can expect that during the period when volatility jumps occur more frequently, this model can generate more accurate forecasts than the HAR-RV model. However, most of time, jump components are zero or are not significant. In these situations, the incorporation of a jump component into a volatility model may lead to overfitting, a situation in which the use of irrelevant predictors may improve the in-sample fitting but lead to poor out-of-sample forecasting performance [25]. Hence, the simple HAR-RV may have greater forecasting accuracy than HAR-RV-J when jumps do not exist, although the latter is more flexible. Therefore, we may have a better forecasting outcome if we switch between different predictive models over time rather than use a single model. In this study, we use DMA to combine individual model forecasts based on their predictive records, and make some improvements on the DMA method, thus becomes the ADMA. Different from DMA method, where models with a history of good predictions receive large weights in the combined future forecasts; ADMA method also diminishes the weights of the models with a history of the poorest predictions. Also the ADMA and DMA method is compared with other popular strategies such as Bayesian model averaging (BMA) and mean forecast combinations (MFC).

Third, we consider the effect of the margin-trading, overnight returns and non-trading days by adding models which put these effect into consideration. Since there are various reasons that affect the performance of the volatility forecasting, it is reasonable to add different types of models as many as possible. We compare the volatility forecasting results of different model sets and prove the non-ignorable effect of the margin-trading, overnight returns and non-trading days.

The remainder of this paper is organized as follows. In Section 2, we briefly describe the popular realized volatility models and explain the methodology used by the DMA and ADMA to combine the forecasts of different models. Section 3 contains the data description and some preliminary analysis. Section 4 reports the main empirical findings. In the last section we present our conclusions.

2. Methodology

In this section, we briefly describe popular realized volatility models and then discuss how we use a dynamic model averaging (DMA) approach and adjusted dynamic model averaging (ADMA) approach to combine the forecasts of the individual models.

2.1. Realized Volatility Measure

In their pioneering work, Andersen and Bollerslev [4] propose using realized volatility (RV) as a proxy for integrated variance. For a specific business day t , the realized volatility can be calculated as the sum of the squared intraday returns $r_{t,j}$:

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

where $1/M$ is the given sampling frequency.

2.2. Modeling Realized Volatility

As for the forecasting model, we start from one of the classic forecasting model proposed by Anderson et al. [6]. In this classic HAR-RV-J model, the decomposition is implemented using the Bi-Power Variation (BPV) measure proposed by Barndorff-Nielsen and Shephard [12], which enables a consistent estimate of the continuous variation in the presence of jumps. The HAR-RV-J model adds an explanatory variable to the standard HAR model introduced by Corsi [13], which has arguably become the most popular model for realized volatility forecasting. Furthermore, Anderson et al. [6] proposed the HAR-RV-CJ model in order to measure the jump process more accurately.

The Corsi's classic model HAR-RV defines as followed:

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + u_t \quad (2)$$

where $RV_{t-j|t-h} = \frac{1}{h+1-j} \sum_{i=j}^h RV_{t-i}$, with $j \leq h$.

So, the HAR-RV-J model comes with the expression:

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + \beta_J J_{t-1} + u_t \quad (3)$$

distributed random variable.
and HAR-RV-CJ model is defined as followed:

where $J_t \equiv \max[RV_t - BPV_t, 0]$, and $BPV_t = \mu_1^{-2} \sum_{i=1}^{M-1} |r_{t,i}| |r_{t,i+1}|$,

with $\mu_1 = \sqrt{2/\pi} = \mathbb{E}(|Z|)$, and Z a standard normally

$$RV_t = \beta_0 + \beta_1 CSP_{t-1} + \beta_2 CSP_{t-1|t-5} + \beta_3 CSP_{t-1|t-22} + \beta_{J,d} CJ_{t-1,d} + \beta_{J,w} CJ_{t-1,w} + \beta_{J,m} CJ_{t-1,m} + u_t \quad (4)$$

where $CJ_t = I(Z_t > \Phi_\alpha) \cdot [RV_t - BPV_t]$, $CSP_t = I(Z_t \leq \Phi_\alpha) \cdot RV_t + I(Z_t > \Phi_\alpha) \cdot BPV_t$.

To capture the role of the ‘‘leverage effect’’ in volatility forecasting, Patton and Sheppard [17] present a series of models using signed realized measures. The first model decomposes the daily RV into two semi-variances (HAR-RV-RS-I),

$$RV_t = \beta_0 + \beta_d^+ RS_{t-1}^+ + \beta_d^- RS_{t-1}^- + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + u_t \quad (5)$$

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

Then, their second model (HAR-RV-RS-II) adds another explanatory variable that interacts the lagged realized variance with an indicator for negative daily returns,

$$RV_t = \beta_0 + \beta_d^+ RS_{t-1}^+ + \beta_d^- RS_{t-1}^- + \gamma RV_{d,t-1} I(r_t < 0) + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + u_t \quad (6)$$

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

$$RV_t = \beta_0 + \beta_{j,d} SJ_{t-1} + \beta_{bpv,d} BPV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + u_t \quad (7)$$

where $SJ_{t-1} = RS_{t-1}^+ - RS_{t-1}^-$.

The fourth model decomposes the signed jump into the positive and the negative ones (HAR-RV-SJ-II):

$$RV_t = \beta_0 + \beta_{j,d}^+ SJ_{t-1}^+ + \beta_{j,d}^- SJ_{t-1}^- + \beta_{bpv,d} BPV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + u_t \quad (8)$$

where $SJ_{t-1}^+ = SJ_{t-1} I(SJ_{t-1} > 0)$ and $SJ_{t-1}^- = SJ_{t-1} I(SJ_{t-1} < 0)$.

Table 1 gives a short summary of the classic RV-based models used in this paper.

Table 1. Typology of the Classic RV-based model specifications.

Model number	Model name	Reference
Model 1	HAR-RV	Corsi (2009)
Model 2	HAR-RV-J	Andersen et al. (2007)
Model 3	HAR-RV-CJ	Andersen et al. (2007)
Model 4	HAR-RV-TCJ	Corsi et al. (2010)
Model 5	HAR-RV-RS-I	Patton and Sheppard (2013)
Model 6	HAR-RV-RS-II	Patton and Sheppard (2013)
Model 7	HAR-RV-SJ-I	Patton and Sheppard (2013)
Model 8	HAR-RV-SJ-II	Patton and Sheppard (2013)

In the real circumstances, a variety of events can occur during non-trading periods, all of which influence the movement of asset prices. The fluctuations at the opening of the markets are the results of the accumulation of information absorption during

market close periods, and they are bound to have a crucial impact on the market during the trading periods [26]; thus, overnight returns are seen as a key element in volatility forecasting. The HAR-RV-J-ONI model defines as following:

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + \beta_J J_{t-1} + \beta_{OD} ONI_{t-1} + u_t \quad (9)$$

where ONI denotes the absolute value of the overnight return, and $J_t \equiv \max[RV_t - BPV_t, 0]$, $BPV_t = \mu_1^{-2} \sum_{i=1}^{M-1} |r_{t,i}| |r_{t,i+1}|$.

Another important factor is non-trading days. So we suggest a simple extension of RV-based model, the NT-HAR model

(non-trading days HAR), which allows the autoregressive coefficient to be dependent on whether a non-trading period occurs between two observations. The HAR-RV-J-NT model defines as following:

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \gamma_0 I[\cdot] + \gamma_1 I[\cdot] RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + \beta_J J_{t-1} + u_t \quad (10)$$

$I[\cdot]$ denotes an indicator function, which is equal to 1 if the calendar time difference between price t and $t-1$ is equal to or greater than 3; otherwise, $I[\cdot]$ equals 0.

Moreover, recently some scholars also consider margin trading constraints as an important factor which affects the volatility forecasting. In the studies where attention had been focused on the effect of short sale constraints, Figlewski [27-28] first used the average short interest of a stock during a particular period to measure the effects of restrictions on short sales. They use the average short interest as percentage of existing shares to construct the dependent variable which measures the effects of restrictions on short sales. Moreover, Boehmer, Ekkehart, and Julie Wu [29-31] proposed a common method of measuring a stock's daily shorting flow. They used a stock's daily shares sold short scaled by its daily share trading volume to measure the daily shorting flow. This standardization makes shorting activity comparable across stocks with different trading volume. These two methods can be applied to the measurement of the effect of short sale constraints on a stock index where the particular stock's daily shares sold short and its' daily share trading volume and outstanding shares can be replaced by the stock index's daily shares sold short and its' daily share trading volume and outstanding shares. This standardization makes the power of the short sale constraints on the stock index comparable across days with different trading volume or outstanding shares. In summary, short interest ratio (SIR) and days to cover (DTC) are two of the most popular proxies for the power of short selling constraints among numerous researches. The calculation formulas for SIR and DTC are listed as follows:

$$SIR = \frac{\text{short interest}}{\text{outstanding shares}} \times 10^4 \quad (11)$$

$$DTC = \frac{\text{short volume}}{\text{trading volume}} \times 10^4 \quad (12)$$

Similarly, we can use the format of these proxies to construct proxies for the power of margin buying constraints alone. For example, we can construct new measures margin buying ratio typeI (MBRI) and margin buying ratio typeII (MBRII) based on the calculation formula of SIR and DTC. And the formula for MBRI and MBRII are listed below:

$$MBRI = \frac{\text{margin buying interest}}{\text{outstanding shares}} \times 10^4 \quad (13)$$

$$MBRII = \frac{\text{margin buying volume}}{\text{trading volume}} \times 10^4 \quad (14)$$

However, these proxies consider only the effect of short sale constraints or margin buying constraints alone. As it has been proved that margin buying constraints lead to undervaluation, it is natural to construct a new short measure that considers both short selling and margin buying constraints. Rui Li et al. [38] propose a new measure, short-margin trading ratio (SMTR), to predict future returns. The SMTR measure is more appropriate than SIR, as it considers the effect of both short selling and margin buying constraints. The calculation formula for SMTR is given as follows:

$$SMTR = \frac{\text{short - selling balance}}{\text{short - selling balance} + \text{margin - buying balance}} \quad (15)$$

We now have these proxies for the two aspects of the effect of margin-trading constraints: effect of short selling constraints and effect of margin buying constraints. We use these proxies to modify the classic RV-based models so that we can verify the effect of short selling and margin buying constraints. For the concision of the article, we only presented the models which will be used in the empirical part. The main idea is to add a variable which measures the power of margin trading constraints into the classic RV-based models. So we present these models' specification below:

HAR-RV-MT model:

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + \beta_{MT} SMTR_{t-1} + u_t \quad (16)$$

HAR-RV-MT-J model:

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + \beta_J J_{t-1} + \beta_{MT} SMTR_{t-1} + u_t \quad (17)$$

HAR-RV-RS-MT-I model:

$$RV_t = \beta_0 + \beta_d^+ RS_{t-1}^+ + \beta_d^- RS_{t-1}^- + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + \beta_{MT} SMTR_{t-1} + u_t \tag{18}$$

HAR-RV-SJ-MT-I model:

$$RV_t = \beta_0 + \beta_{j,d} SJ_{t-1} + \beta_{bpv,d} BPV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + \beta_{MT} SMTR_{t-1} + u_t \tag{19}$$

And Table 2 summarizes the modified RV-based models which consider the effect of overnight returns, non-trading days and margin trading constraints.

Table 2. Typology of the Modified RV-based model specifications.

Model number	Model name	Reference
Model 9	HAR-RV-ONI	Tseng et al. (2009)
Model 10	HAR-RV-ONI-J	Tseng et al. (2009)
Model 11	HAR-RV-NT	Lócsa et al. (2017)
Model 12	HAR-RV-NT-J	Lócsa et al. (2017)
Model 13	HAR-RV-MT	
Model 14	HAR-RV-MT-J	See section 2.2 part of the paper
Model 15	HAR-RV-RS-MT-I	
Model 16	HAR-RV-SJ-MT-I	

2.3. Dynamic Model Averaging and Adjusted Form

The realized volatility models expressed in the section 2.2 are actually regressions with constant coefficient models. In this study, we extend the models to include time-varying parameter specifications. Suppose that we have K number of models, each of which takes a different set of predictors. These time-varying parameter RV models can be briefly written as

$$RV_t = x_{t-1}^{(k)} \beta_t^{(k)} + \varepsilon_t \tag{20}$$

$$\beta_t^{(k)} = \beta_{t-1}^{(k)} + \eta_t^{(k)} \tag{21}$$

where the vector $x_{t-1}^{(k)}$ is the set of predictors of model k including the intercept. For instance, in the standard HAR-RV model, $x_{t-1}^{(k)}$ contains an intercept, lagged daily RV, weekly RV, and monthly RV; the vector $\beta_t^{(k)}$ represents the time-varying regression coefficients; $\varepsilon_t \sim N(0, H_t)$ and $\eta_t \sim N(0, Q_t)$. The errors ε_t and η_t are assumed to be mutually independent at all of the leads and lags.

As for the method for estimating H_t , we choose the Exponentially Weighted Moving Average (EWMA) method:

$$H_t^{(k)} = \frac{1-\kappa}{1-\kappa^{j-1}} \sum_{j=1}^{t-1} \kappa^j (RV_{t-j} - x_{t-j-1}^{(k)} \beta_{t-j}^{(k)})^2 \tag{22}$$

where the decay factor κ is set to be 0.94, according to the suggestion of Riskmetrics.

Let $L_t \in \{1, 2, \dots, K\}$ denote which model is used to make a forecast at time t, $B_t = (\beta_t^{(1)}, \dots, \beta_t^{(K)})'$, and $RV^t = (RV_1, \dots, RV_t)'$. In the procedure of DMA, we will let all of these different models hold at each time and then perform model averaging. Specifically, when forecasting the realized

volatility at time t, we use all of the information through time t-1 and for $k = 1, 2, \dots, K$. DMA involves in calculating the probability of being model k. Then, these probabilities are used to average the forecasts across the different models.

The adjusted dynamic model averaging (ADMA) is quite similar with DMA, the main difference between these two methods lies in the use of the probability of each individual model. In the DMA method, all the probability of each particular model is used to average the forecasting results. But in the ADMA method, the least and second least probability of particular model k_1 and k_2 are excluded from the process of averaging the forecasts across the different models. As a special case, we also present the DMS method which only selects a single model with the highest value for $\Pr(L_t = k | RV^{t-1})$. See the details on how to estimate DMA and ADMA with TVP models in Appendix A.

3. Data and Preliminary Analysis

In this study, we use 5-min price data from the CSI 300 index for the April 1, 2010 to June 17, 2015 period. The data are for the trading time of each business day between 9:30:00 and 15:00:00. After removing days with shortened trading sessions or too few transactions, we obtain high-frequency data for 1266 business days. All of the price data are taken from the high-frequency database of the stock market established by the China Centre for Economic Research.

The start of the period is the time when the CSRC began to allow short selling and margin buying, and the end of the period is the stock market disaster that began in the middle of 2015 in China's stock market. We use the CSI300 as our sample, as it reflects the overall trend in the two markets. The index sample is drawn from the Shanghai and Shenzhen stock markets and covers most of the circulation of market value. Component stocks in the CSI300 are mainstream investment stocks with good market representation, high liquidity, and active trading, and thus they reflect the return of mainstream investment in the market. Moreover, over 95% of the component stocks in the CSI300 are allowed to participate in short selling and margin buying. Thus, the CSI300 also contains information about the effect of margin trading constraints, which is the main concern of this study. We also retrieve data on short selling and margin buying during the study period from the China Stock Market and Accounting Research database provided by the GuoTaiAn Company.

The panel A of Table 3 presents the descriptive statistics of the RV measures used in this study. All of the series are right-skewed and display positive kurtosis, which suggests that they have non-Gaussian distributions. Moreover, the logarithmic standard deviations are generally much closer to being normally distributed than are the raw realized volatility

series [7, 32]. Hence, from a modeling perspective, the logarithmic realized volatility is more amenable to the use of standard time series procedures. So we report the descriptive statistics of the logarithmic form of RV in panel B and use the

logarithmic form of RV to construct our predicting models in the in-sample forecasting. And we provide descriptive statistics of our proxies for the power of the short selling and margin buying constraints in panel C.

Table 3. Summary Statistics.

	Mean	Median	Maximum	Minimum	Std. Dev.	Skewness	Kurtosis
Panel A Summary statistics for realized volatility (normal form)							
RV	1.437	0.941	27.603	0.115	1.883	7.311	82.559
BPV	1.284	0.819	31.759	0.108	1.969	9.189	122.169
J	0.180	0.104	3.249	0	0.261	4.086	30.631
RS+	0.734	0.487	10.960	0.044	0.869	5.154	44.817
RS-	0.699	0.420	17.396	0.039	1.122	8.389	102.327
SJ	0.034	0.031	3.943	-7.767	0.698	-2.998	36.551
SJ+	0.205	0.031	3.943	0	0.386	3.769	23.403
SJ-	-0.171	0	0	-7.767	0.517	-8.262	97.156
Panel B Summary statistics for realized volatility (logarithmic form)							
Log RV	0.021	-0.027	1.441	-0.941	0.329	0.586	3.887
Log BPV	-0.055	-0.087	1.502	-0.965	0.337	0.689	4.204
Log J	0.064	0.043	0.628	0	0.075	2.300	10.851
Log RS+	-0.296	-0.312	1.039	-1.354	0.359	0.294	3.223
Log RS-	-0.342	-0.376	1.240	-1.407	0.364	0.571	4.025
Log SJ	0.045	0.039	1.028	-0.911	0.303	0.097	3.062
Log SJ+	0.143	0.039	1.028	0	0.197	1.519	4.844
Log SJ-	-0.097	0	0	-0.911	0.159	-1.877	6.244
Panel C Summary statistics for margin trading constraints proxy							
MBRI	0.057	0.011	0.948	3.24e-04	0.684	1.348	4.087
MBRII	0.093	0.059	0.390	0.041	0.086	0.798	2.580
SIR	0.002	5.03e-05	0.036	1.28e-05	4.37e-05	0.430	2.981
DTC	0.015	0.015	0.056	0	0.014	0.425	2.026
SMTR	0.013	0.009	0.043	1.60e-04	8.56e-03	1.079	3.843

4. Empirical Results

4.1. Forecasting Performances of Individual Models

Compared with the in-sample performance, the out-of-sample performance of a model (i.e., its predictive ability) is more important to market participants, since they are more concerned about the model's ability to improve their future performance than its ability to analyze past patterns. So in this paper, The rolling window method is applied to forecast volatility for the period from April 1, 2010 to April 23, 2012 (500 data points). And the size of the rolling window is also fixed at the level of 500 data points. To evaluate the forecasting accuracy of the volatility models, we use the following loss functions and accuracy statistics:

$$MSE = M^{-1} \sum_{m=H+1}^M (RV_m - \hat{\sigma}_m^2)^2 \quad (23)$$

$$MSD = M^{-1} \sum_{m=H+1}^M (RV_m^{1/2} - \hat{\sigma}_m)^2 \quad (24)$$

$$MAE = M^{-1} \sum_{m=H+1}^{H+M} |RV_m - \hat{\sigma}_m^2| \quad (25)$$

$$MAD = M^{-1} \sum_{m=H+1}^{H+M} |RV_m^{1/2} - \hat{\sigma}_m| \quad (26)$$

where RV_m is the actual RV; $\hat{\sigma}_m^2$ is the volatility forecasts, and M is the number of forecasting data points.

To check whether the differences of loss functions are significant, we assess the statistical significance of differences in forecasting losses using the model confidence set (MCS) developed by Hansen et al. [33]. MCS chooses a subset of models containing all possible optimal models from the initial model set. Based on Hansen et al. [33], we use the confidence level of 75% which indicates the models that exclude from our subset are significantly less accurate than the models in the MCS.

To investigate the effects of time-variation on out-of-sample performance, we consider two types of predictive regressions depending on the different pattern of parameters. The first is the constant coefficient (CC) model assuming that the regressive parameters do not change over time. The second is the time-varying parameter (TVP) model which can capture the change of predictive relationship at each point of time.

We compare the forecasting performances of two types of predictive regressions with CC and TVP specifications. Table 4 shows the forecasting results of individual models of original RV-based models (type 1). We can see that TVP models always generate volatility forecasts that have lower loss functions than CC models, implying the greater forecasting accuracy. Volatility forecasts from CC models have the highest loss functions, suggesting the worst forecasting performance. The MCS test results indicate that under the criterion of MSE, the performances of two types of

models are not significantly different. However, under the other three loss criteria CC models are excluded in most cases regardless of which volatility model specification is used. Therefore, we can conclude that relative performances of two

types of models depend on the use of loss criteria but TVP models are better choices than CC ones. These results highlight the importance of allowing for time-variation in parameters of predictive regressions.

Table 4. Forecasting performances of individual models (type 1) evaluated by loss functions.

	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8
MSE								
CC	3.673	3.666	3.721	3.704	3.686	3.712	3.659	3.968
TVP	3.022	3.015	3.027	3.033	3.016	3.019	3.007	3.144
MSD								
CC	0.706	0.704	0.705	0.704	0.708	0.710	0.706	0.724
TVP	0.633	0.631	0.630	0.633	0.635	0.644	0.640	0.658
MAE								
CC	0.159	0.158	0.158	0.158	0.159	0.159	0.158	0.169
TVP	0.116	0.114	0.115	0.115	0.116	0.116	0.115	0.116
MAD								
CC	0.244	0.244	0.245	0.246	0.245	0.246	0.245	0.247
TVP	0.207	0.206	0.206	0.208	0.209	0.211	0.208	0.209

Notes: This table provides the out-of-sample forecasting results of 8 individual RV-based classic models evaluated by four loss functions. The numbers in bold denote that the corresponding model has the lowest loss function under a specific criterion.

Similarly, table 5 shows the forecasting results of individual models of modified RV-based models considering margin trading, overnight returns and non-trading days (type 2) evaluated by loss functions. Not surprisingly, no matter which loss function is used, the time-varying (TVP) model is always

superior to the fixed parameter (CC) model, and the forecasting accuracy of the two types of model is obviously different under different loss functions, which further shows the advantage of the time-varying model in forecasting future volatility.

Table 5. Forecasting performances of individual models (type 2) evaluated by loss functions.

	Model 9	Model 10	Model 11	Model 12	Model 13	Model 14	Model 15	Model 16
MSE								
CC	0.523	0.529	0.544	0.507	0.246	0.237	0.244	0.241
TVP	0.511	0.505	0.527	0.496	0.219	0.213	0.218	0.212
MSD								
CC	0.406	0.417	0.421	0.394	0.234	0.235	0.236	0.237
TVP	0.344	0.349	0.355	0.346	0.216	0.218	0.218	0.219
MAE								
CC	0.085	0.086	0.081	0.083	0.033	0.034	0.035	0.035
TVP	0.048	0.045	0.049	0.042	0.021	0.022	0.022	0.024
MAD								
CC	0.207	0.206	0.211	0.210	0.137	0.138	0.139	0.139
TVP	0.122	0.116	0.116	0.117	0.092	0.091	0.104	0.103

Notes: This table provides the out-of-sample forecasting results of 8 RV-based modified models evaluated by four loss functions. The numbers in bold denote that the corresponding model has the lowest loss function under a specific criterion.

4.2. Forecasting Performance of DMA and ADMA

It has been well documented in the literature that the predictive ability of a single model is quite unstable but changes over time [52]. For this consideration, dynamic model averaging (DMA) and adjusted dynamic model averaging (ADMA) are imposed on individual models. In order to further investigate how well ADMA and DMA performs out-of-sample, we also consider some alternative strategies. These strategies including Dynamic model selection (DMS), Mean forecast combination (MFC) and Trimmed mean combination (TMC) are imposed over CC and TVP models. They have some different properties from DMA. DMS is actually a special case of DMA that chooses the forecasts from the model with the highest posterior probability at each point of time. Mean forecast combination (MFC) strategy uses the simple equal-weighted average of forecasts from individual

models under consideration. Trimmed mean combination (TMC) method uses the equal weighted average of forecasts from individual models after trimming the one with the worst past performance. Imposing these methods over all CC and TVP specifications, we have nine different strategies:

- (1) DMS over TVP models (DMS-TVP).
- (2) DMA over constant coefficient volatility models (DMA-CC).
- (3) DMA over time-varying parameter volatility models (DMA-TVP).
- (4) MFC over constant coefficient volatility models (DMA-CC).
- (5) MFC over time-varying parameter volatility models (DMA-TVP).
- (6) TMC over constant coefficient volatility models (TMC-CC).
- (7) TMC over time-varying parameter volatility models

- (TMC-TVP).
 (8) ADMA over constant coefficient volatility models (ADMA-CC).
 (9) ADMA over time-varying parameter volatility models (ADMA-TVP).

In summary, we consider a total of 9 model averaging and forecast combination strategies including the main strategy of ADMA-TVP. Tables 6 and 7 report the forecasting performances of above mentioned strategies evaluated by four loss functions, as well as the MCS p-values in different model set situation. We first use the model set 1 which only includes the fundamental RV-based models presented in Table 1, then we add the modified RV-based models listed in Table 2 into model set 2 to see whether the results are the same. Some interesting patterns can be found.

First, regardless of the rules that give weight to each model, the prediction performance of multi-model realized volatility prediction model is better than that of single model in most cases, which indicates that the multi-model realized volatility prediction model can effectively improve the prediction accuracy and stability.

Second, the realized volatility prediction model based on the adjusted dynamic averaging (ADMA-TVP) has the best prediction accuracy. Although the improvement on the forecasting performance of the ADMA-TVP model is not obvious compared with the DMA-TVP model, we can still draw the conclusion that the prediction accuracy of the future volatility can be improved effectively by removing the two models with the lowest probability at each time point from the DMA-TVP model.

Third, as noted in Wang et al. [1], our empirical results also show that the performance of the realized volatility model with time-varying parameters based on various rules is not statistically significant. Both DMS and DMA methods can eliminate the instability of a single model in prediction, but there is no clear answer to the specific choice of the rules to give weight to each model. The adjusted dynamic averaging (ADMA) rule presented in this article outperforms DMS and DMA in some loss functions, which leaves us an alternative choice in choosing the weighting rule of the model.

Table 6. Forecasting performances of DMA and alternative strategies evaluated by loss functions (Model Set 1).

	MSE	MSD	MAE	MAD
DMS-TVP	3.237	0.652	0.116	0.211
DMA-CC	3.478	0.715	0.163	0.246
DMA-TVP	3.234	0.647	0.115	0.207
MFC-CC	3.515	0.723	0.158	0.249
MFC-TVP	3.242	0.646	0.115	0.211
TMC-CC	3.510	0.719	0.157	0.248
TMC-TVP	3.230	0.645	0.115	0.210
ADMA-CC	3.455	0.717	0.160	0.246
ADMA-TVP	3.231	0.645	0.116	0.208

Notes: This table provides the out-of-sample forecasting results of models using ADMA, DMA and other strategies evaluated by four loss functions under model set 1. The numbers in bold denote that the corresponding model has the lowest loss function under a specific criterion. The numbers with underlines denote that the corresponding models are included in MCS for the confidence of 75% from Hansen et al. (2011) model encompassing test.

Table 7. Forecasting performances of DMA and alternative strategies evaluated by loss functions (Model Set 2).

	MSE	MSD	MAE	MAD
DMS-TVP	3.016	0.637	0.116	0.208
DMA-CC	3.678	0.705	0.159	0.245
DMA-TVP	3.009	0.632	0.115	0.207
MFC-CC	3.724	0.708	0.160	0.245
MFC-TVP	3.035	0.638	0.115	0.208
TMC-CC	3.671	0.706	0.158	0.245
TMC-TVP	3.008	0.633	0.115	0.209
ADMA-CC	3.662	0.703	0.158	0.244
ADMA-TVP	3.005	0.633	0.116	0.206

Notes: This table provides the out-of-sample forecasting results of models using ADMA, DMA and other strategies evaluated by four loss functions under model set 2. The numbers in bold denote that the corresponding model has the lowest loss function under a specific criterion. The numbers with underlines denote that the corresponding models are included in MCS for the confidence of 75% from Hansen et al. (2011) model encompassing test.

4.3. Robustness Test

For the dynamic average time-varying realized volatility prediction model based on DMA and ADMA, there are two parameters: forgetting factor α and κ , which will affect the iterative speed in the dynamic model averaging process, and thus affect the weight given by each model at different time points. In the robustness test of this section, we use different sum of forgetting factors to study the robustness of prediction model based on DMA and ADMA for future volatility prediction. In the empirical part, we set $\alpha = 0.995$ and $\kappa = 0.95$, and here we will also refer to the Wang's article [1] to set the forgetting factor α to 0.95, thus giving the model a faster iteration speed as a whole. In addition, we set another forgetting factor κ to 0.90, and put the DMA-TVP and ADMA-TVP models based on these different forgetting factor parameters together with the original DMA-TVP and ADMA-TVP models for MCS test. The results shown in Table 8 show that the prediction accuracy of DMA and AMDA with different degrees of forgetting is fairly close under any of the loss function evaluation criteria. The results of MCS test also show that the consistency of prediction accuracy of these DMA and ADMA models based on different parameters is the same statistically. Of course, an obvious fact is that the prediction accuracy of these DMA and ADMA models based on different parameters is higher than that of most single models, which reflects the robustness of the multi-model realized volatility prediction model based on dynamic averaging method in the prediction effect.

Table 8. Forecasting performances of DMA for alternative priors.

	MSE	MSD	MAE	MAD
DMA ($\alpha = 0.995, \kappa = 0.95$)	3.009	0.632	0.115	0.207
DMA ($\alpha = 0.95, \kappa = 0.95$)	3.010	0.630	0.116	0.206
DMA ($\alpha = 0.995, \kappa = 0.90$)	3.009	0.633	0.117	0.206
DMA ($\alpha = 0.95, \kappa = 0.90$)	3.012	0.632	0.116	0.209
ADM ($\alpha = 0.995, \kappa = 0.95$)	3.005	0.633	0.116	0.206
ADMA ($\alpha = 0.95, \kappa = 0.95$)	3.007	0.635	0.115	0.207
ADM ($\alpha = 0.995, \kappa = 0.90$)	3.006	0.632	0.114	0.208
ADMA ($\alpha = 0.95, \kappa = 0.90$)	3.006	0.632	0.115	0.208

5. Conclusions

Intra-day high-frequency data is now widely available, and since the pioneer work of Andersen and Bollerslev [3], the forecasting of realized volatility (RV) has become an important aspect of research. In this study, we forecast the realized volatility of the CSI 300 index. Different from previous studies that use a single model with constant coefficients, we use consider time-variation in parameters due to structural breaks. Moreover, as the forecasting accuracy of each individual model is very unstable, we use a adjusted dynamic model averaging (ADMA) approach based on the DMA method to address this issue of model uncertainty. ADMA addresses the problem of the model specification changing over time by combining the forecasts generated from different models and has a better forecasting performances theoretically. Our out-of-sample results indicate that time-varying parameter (TVP) models can generate more accurate forecasts than constant coefficient models in China’s stock market. ADMA forecasts also have greater forecasting accuracy than DMA method and some traditional forecast combination strategies. These findings indicate the importance of considering parameter change and model specification change in volatility forecasting.

Appendix

Parameter Estimation of TVP Models

In order to reduce the calculation burden, we follow Raftery et al. [22] in using approximations based on two forgetting factors α and κ . Below we describe how these forgetting factors work in parameter estimation. For the covariance matrices H_t and Q_t , Kalman filtering is used to estimate and forecast the volatility based on the following two equations:

$$\begin{aligned} \beta_{t-1}^{(k)} | RV^{t-1} &\sim N(\hat{\beta}_{t-1}^{(k)}, \sum_{t-1|t-1}^{(k)}) \\ \beta_t^{(k)} | RV^{t-1} &\sim N(\hat{\beta}_{t-1}^{(k)}, \sum_{t|t-1}^{(k)}) \end{aligned} \tag{27}$$

where $\sum_{t|t-1}^{(k)} = \sum_{t-1|t-1}^{(k)} + Q_t^{(k)}$.

Starting at $t = 0$, Kalman filtering updates these formulae and makes a prediction based on the predictive distribution,

$$RV_t | RV^{t-1} \sim N(x_{t-1} \hat{\beta}_{t-1}^{(k)}, H_t^{(k)} + x_{t-1}^{(k)} \sum_{t|t-1}^{(k)} x_{t-1}^{(k)}) \tag{28}$$

Here, Raftery et al. [22] use the forgetting factor method based on the equation,

$$\sum_{t|t-1}^{(k)} = \frac{1}{\lambda} \sum_{t-1|t-1}^{(k)} \tag{29}$$

This approach has a long history in the state space literature. The use of a “forgetting factor” in this specification implies that the observations in past j periods are assigned the weight of λ^j . As a special case, $\lambda = 1$ corresponds to the constant coefficient model.

After obtaining volatility forecasts based on the individual models, one needs to combine these forecasts using DMA and

ADMA. For the DMA method, let $\pi_{t|s,k} = \Pr(L_t = k | y^s)$, then the new recursions required by DMA are $\pi_{t|t-1,k}$ and $\pi_{t|t,k}$. DMA averages the forecasts of the different models using $\pi_{t|t-1,k}$ as weights for $k = 1, \dots, K$ and $t = 1, \dots, T$. For instance, the DMA forecasts can be defined as,

$$E(RV_t | RV^{t-1}) = \sum_{k=1}^K \pi_{t|t-1,k} x_t^{(k)} \hat{\beta}_{t-1}^{(k)} \tag{30}$$

where $\hat{\beta}_{t-1}^{(k)}$ are the Kalman filter estimates of the regression coefficients at time $t-1$.

Raftery et al. [22] use the following equation to describe the relation between $\pi_{t|t-1,k}$ and $\pi_{t|t,k}$:

$$\pi_{t|t-1,k} = \frac{\pi_{t-1|t-1,k}^\alpha}{\sum_{l=1}^K \pi_{t-1|t-1,l}^\alpha} \tag{31}$$

In this paper, we use a simpler evaluation by updating the equation as follows:

$$\pi_{t|t,k} = \frac{\pi_{t-1|t-1,k} p_k(RV_t | RV^{t-1})}{\sum_{l=1}^K \pi_{t-1|t-1,l} p_l(RV_t | RV^{t-1})} \tag{32}$$

where $p_k(y_t | y^{t-1})$ is the predictive density of model k evaluated at RV_t . This implies that the weight of model k at time t is

$$\begin{aligned} \pi_{t|t-1,k} &\propto [\pi_{t-1|t-2,k} p_k(RV_{t-1} | RV^{t-2})]^\alpha \\ &= \prod_{i=1}^{t-1} [p_k(RV_{t-i} | RV^{t-i-1})]^\alpha \end{aligned} \tag{33}$$

Therefore, a model that has had a better forecasting performance in the past, as measured by the predictive density $p_k(RV_{t-i} | RV^{t-i-1})$, will receive more weight at time t . The relative importance of past forecasting performance is controlled by the forgetting factor α , which has an exponential decay rate of α^i .

The ADMA method is similar with DMA, the only difference is after calculating the weight of each model in the model set, we let the least two weight of the model become 0. Which means, a model that has the worst and the second worst forecasting performance in the past, as measured by the predictive density $p_k(RV_{t-i} | RV^{t-i-1})$, will receive zero weight at time t .

References

[1] Wang et al. (2016). Forecasting realized volatility in a changing world: A dynamic model averaging approach. *Journal of Banking & Finance*, 64, 136-149.

- [2] Amihud, Y., & Mendelson, H. (2006). Asset pricing and the bid-ask spread. *Journal of Financial Economics*, 17 (2), 223-249.
- [3] Amihud, Y., & Mendelson, H., Wood R., (1990). Liquidity and the 1987 stock market crash. *Journal of Portfolio Management*, 44, 65-69.
- [4] Andersen, T. G., Bollerslev, T., (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39 (4), 885-905.
- [5] Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., (2003). Modeling and forecasting realized volatility. *Econometrica*, 71 (2), 579-625.
- [6] Andersen, T. G., Bollerslev, T., Diebold, F. X., (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *Review of Economics and Statistics*, 89 (4), 701-720.
- [7] Andersen, T. G., Bollerslev, T., Diebold, F. X., H. Ebens, (2001). The Distribution of Realized Stock Return Volatility. *Journal of Financial Economics*, 61, 43-76.
- [8] Andersen, T. G., Bollerslev, T., Meddahi, N., (2011). Realized volatility forecasting and market microstructure noise. *Journal of Econometrics*, 160 (1), 220-234.
- [9] Asquith, Paul & Pathak, Parag A. & Ritter, Jay R., (2005), Short interest, institutional ownership, and stock returns, *Journal of Financial Economics*, 78 (2), 243-276.
- [10] Barndorff-Nielsen, O. E., et al. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76 (6), 1481-1536.
- [11] Barndorff-Nielsen, O. E., Shephard, N., (2006). Econometrics of testing for jumps in financial econometrics using bipower variation. *Journal of Financial Econometrics*, 4 (1), 1-30.
- [12] Barndorff-Nielsen, O. E., Shephard, N., (2004b). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2 (1), 1-37.
- [13] Corsi, F., (2004). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7 (2), 174-196.
- [14] Corsi, F., Pirino, D., Reno, R., (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*, 159, 276-288.
- [15] Corsi, F., Reno, R., (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business and Economic Statistics*, 30, 368-380.
- [16] Patton, A. J., (2011). Data-based ranking of realized volatility estimators. *Journal of Econometrics*, 161 (2), 284-303.
- [17] Patton, A. J., Sheppard, K., (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Reviews of Economics and Statistics*, 97 (3), 683-697.
- [18] Ping, Yuan, Li, et al. (2018). Forecasting realized volatility based on the truncated two-scales realized volatility.
- [19] Zhang, L., Mykland, P. A., Ait-Sahalia, Y., (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of American Statistical Association*, 100 (472), 1394-1411.
- [20] Calvet, L. E., Fisher, A. J., (2004). How to forecast long-run volatility: regime switching and the estimation of multifractal processes. *Journal of Financial Econometrics* 2, 49-83.
- [21] Hamilton, J. D., Susmel, R., (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* 64, 307-333.
- [22] Raftery, A. E., Kárny, Miroslav, & Ettlér, P. (2010). Online prediction under model uncertainty via dynamic model averaging: application to a cold rolling mill. *Technometrics*, 52 (1), 52-66.
- [23] Hibon, M., Evgeniou, T., (2004). To combine or not combine: selecting among forecasts and their combinations. *International Journal of Forecasting* 21, 15-24.
- [24] Liu, C., Maheu, J. M., (2009). Forecasting realized volatility: a Bayesian model averaging approach. *Journal of Applied Econometrics* 24, 709-733.
- [25] Scheinkman, J. A. and Xiong, W. (2004) Overconfidence and Speculative Bubbles. *Journal of Political Economy*, 111, 1183-1219.
- [26] Giampiero M. Gallo, Barbara Pacini. (2000). The effects of trading activity on market volatility. *European Journal of Finance*, 6 (2), 163-175.
- [27] Figlewski, S. (1981). The informational effects of restrictions on short sales: some empirical evidence. *Journal of Financial & Quantitative Analysis*, 16 (4), 463-476.
- [28] Figlewski, S., & Webb, G. P. (1993). Options, short sales, and market completeness. *Journal of Finance*, 48 (2), 761-777.
- [29] Boehmer, E., Huszar, Z. R., & Jordan, B. D. (2010). The good news in short interest. *Social Science Electronic Publishing*, 96 (1), 80-97.
- [30] Boehmer, Ekkehart, and Julie Wu, (2010). Short selling and the informational efficiency of prices. Working paper, Texas A&M University.
- [31] Boehmer, E., Huszar, Z. R., Wang, Y., & Zhang, X. (2017). Are shorts equally informed? a global perspective. *Social Science Electronic Publishing*.
- [32] Deo, R., C. Hurvich, and Y. Lu, (2006). Forecasting Realized Volatility Using a Long-Memory Stochastic Volatility Model: Estimation, Prediction and Seasonal Adjustment. *Journal of Econometrics*, 131, 29-58.
- [33] Hansen, P. R., Lunde, A. and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79, 453-97.