Economic Modelling 61 (2017) 12–26

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.elsevier.com/locate/econmod

Forecasting the realized range-based volatility using dynamic model averaging approach

Jing Liu[a], Yu Wei[a], Feng Ma[a,*], M.I.M. Wahab[b]

[a] School of Economics & Management, Southwest Jiaotong University, Chengdu, Sichuan, PR China610031
[b] Department of Mechanical and Industrial Engineering, Ryerson University, 350 Victoria Street, Toronto, ON, Canada5MB 2K3

ARTICLE INFO

Jel Classification:
C22
C52
C55

Keywords:
Volatility forecasting
Realized range-based volatility
Dynamic model averaging
Combined models

ABSTRACT

In this study, we forecast the realized range-based volatility (RRV) using the heterogeneous autoregressive realized range-based volatility (HAR-RRV) model and its various extensions, which are called HAR-RRV-type models. We first consider the time-varying property of those models' parameters using the dynamic model averaging (DMA) approach and evaluate the forecasting performance of three types: individual HAR-RRV-type models, combined models with constant weights, and combined models with time-varying weights. Our out-of-sample empirical results show that combined models with time-varying weights can not only generate more accurate forecasts, but also beat individual models and combined models with constant weights.

1. Introduction

The volatility of financial asset and commodity price representing their price uncertainty has a huge strategic importance in risk management, derivative pricing, and portfolio selection. For example, crude oil is traded in the global market and its price uncertainty has significant effect on economic growth and financial market all around the world (see, e.g., Hamilton, 1983; Kilian, 2006; Aloui and Jammazi, 2009; Kilian and Park, 2009). Therefore, modelling and forecasting volatility of asset price are crucial to financial market participants and policy makers.

A large number of studies focus on modelling and forecasting volatility of financial and commodity markets using low-frequency data (e.g., Agnolucci, 2009; Cheong, 2009; Wei et al., 2010; Mohammadi and Su, 2010; Nomikos and Andriosopoulos, 2012; Charles and Darné, 2014; Efimova and Serletis, 2014; Lean and Smyth, 2015). As high-frequency data carries more information, it can help make better decisions. With the increasing availability of high-frequency data, research on measuring and modelling the volatility based on high-frequency data becomes popular.

For measuring volatility of high-frequency data, Andersen and Bollerslev (1998) propose the realized volatility (RV), which is defined as the sum of non-overlapping squared returns within a fixed time interval. Corsi (2009) proposes a simple heterogeneous autoregressive model of realized volatility (HAR-RV) model. This model is popularly employed in forecasting volatility and shows its outstanding performance in capturing "stylized facts" in financial market, such as long memory and multi-scale behavior of volatility (Andersen et al., 2007; Corsi et al., 2010; Bekaert and Hoerova, 2014; Duong and Swanson, 2015; Bollerslev et al., 2016). Following Corsi (2009), some extensions of HAR-RV model are developed (Andersen et al., 2007; Huang et al., 2016). For example, HAR-RV-J model is proposed by adding the jump component in volatility and HAR-RV-CJ model is developed to investigate the contribution of jumps by decomposing realized volatility into continuous sample path and significant jump components (Andersen et al., 2007). These extensions are mainly based on different decompositions of realized volatility. Moreover, by introducing leverage terms related to negative returns, McAleer and Medeiros (2008) extend HAR-RV model to the leverage HAR-RV (LHAR-RV) model. Recently, some empirical studies show overnight information has a significant impact on volatility and thus can improve the predictability of HAR-type models (e.g., Taylor, 2007; Todorova and Souček, 2014). Overall, HAR-type models perform better than GARCH-class models in capturing the volatility dynamics and are more widely used in high-frequency data (see, e.g., Andersen et al., 2003; Hansen and Lunde, 2005; Koopman et al., 2005; Wei et al., 2010; Çelik and Ergin, 2014; Ma et al., 2015).

Nevertheless, Bandi and Russell (2008) point out that RV could not identify the daily integrated variance of the frictionless equilibrium price in the presence of market microstructure noise. Thus, Martin and

* Corresponding author.
E-mail address: mafeng2016@swjtu.edu.cn (F. Ma).

http://dx.doi.org/10.1016/j.econmod.2016.11.020
Received 1 June 2016; Received in revised form 12 November 2016; Accepted 25 November 2016
Available online 09 December 2016
0264-9993/

Dick (2007) introduce another measure called the realized range-based volatility (RRV), which is calculated by the difference between the largest and lowest prices that are observed during a certain period. The measurement precision of RRV is proved to be up to five times greater than the RV. Since the extremes are obtained from the entire price process, RRV carries more information and also relatively less contaminated by noise than realized volatility at fixed intervals. RRV has already been applied in financial assets and it successfully captures the long-term memory behavior in volatility (e.g., Tseng et al., 2009). Some models, such as HAR-RRV and HAR-RRV-RBV-ONI model, are constructed based on RRV and show good out-of-sample forecasting performances (Tseng et al., 2012). Since RRV is much more efficient and sufficient in volatility modelling and forecasting, we choose RRV to measure the volatility of financial assets.

Although, many individual models, such as GARCH-class models and HAR-type models, have been presented, it has also been well documented that the predictability of an individual model is very unstable and changing over time (e.g., Stock and Watson, 2004). Hence, some studiesdiscuss how to combine a set of forecasts to produce superior composite forecasts (Liu and Maheu, 2008; Santos and Ziegelmann, 2014). The existing literature shows that a combined model performs better than an individual model.

In addition, asset volatility is affected by many uncertain factors, such as economic cycles, political policies, and extreme events, which will lead to frequent structural breaks in the statistical property of volatility (e.g., Granger and Hyung, 2004; Liu and Maheu, 2008). To deal with structural breaks over time in a single model, Raftery et al. (2010) propose the dynamic model averaging (DMA) approach, which allows the model vary with the variables over time. Consequently, DMA is widely implemented to forecast inflation, gold price, oil price, and combine the forecasts (Koop and Korobilis, 2012; Aye et al., 2015; Naser, 2016; Wang et al., 2016). DMA approach combines the generated models dynamically by using two forgetting factors to approximately estimate both the model parameters and model switching probabilities, i.e., DMA successfully avoids arbitrary choices made by the model users. However, there is no study of realized range-based volatility dynamics which are described in a time-varying parameter framework such as dynamic model averaging. In order to fill this gap, we first incorporate DMA approach into RRV framework for forecasting volatility of crude oil futures and the S & P 500 index.

The validity of forecasting models is usually evaluated by various methods including loss functions, mean mixed statistics (MME), superior predictive ability (SPA) test, and advanced model confidence set (MCS) test (Brailsford and Faff, 1996; Hansen, 2005; Hansen et al., 2011). Among them, the MCS test has several attractive advantages. MCS test does not require a benchmark and allows for the possibility of more than one "best" models. Thus, we apply MCS as a main criterion for model evaluation.

To the best of authors' knowledge, this study is the first attempt to incorporate DMA approach into RRV framework. Thus, both the time-varying property of high-frequency volatility and time-varying weights of different models are considered over time. In order to forecast the realized range-based volatility of financial assets, we construct five individual HAR-RRV-type models (HAR-RRV, HAR-RRV-J, HAR-RRV-CJ, LHAR-RRV-CJ, and HAR-RRV-ONI), their combinations with constant weights, and their combinations with time-varying weights by applying the DMA approach. Finally, we assess their predictive abilities by using error statistics, MME test, and MCS test.

The contributions of this research are threefold. First, we model the time-varying volatility and compare the models' forecasting ability in the framework of RRV, since the RRV carries more information than RV. We also provide the performance of the models for forecasting RRV across crude oil futures and the S & P 500 index. Second, individual HAR-RRV-type models are unstable over time. However, there is no study on incorporating time-varying combined models with the framework of RRV. Hence, we employ several combined models with constant weights and combined models time-varying weights obtained from DMA to capture the dynamics of volatilities. Third, our empirical study evaluates three types of models (individual HAR-RRV-type models, combined models with constant weights, and combined models with time-varying weights) in forecasting volatility based on error statistics, mean mixed statistics (MME), and MCS test for 5-min, 10-min, and 15-min high-frequency data. Our empirical studies show that DMA approach performs the best. Especially, DMA shows its strength for forecasting RRV based on 5-min frequency data of oil futures.

The rest of this paper is organized as follows. Section 2 briefly describes the specifications of the volatility measure and five individual models based on RRV. Section 3 presents the high-frequency data. Section 4 discusses the in-sample evaluations and out-of-sample forecasting results. Section 5 concludes the paper.

2. Methodology

The volatility measure and forecasting methodology are introduced in this section. Section 2.1 describes the measure of realized range-based volatility (RRV). Based on RRV, HAR-RRV model and its extensions are specified in Section 2.2. Section 2.3 presents the combined models and Section 2.4 discusses applying DMA approach as a combined model with time-varying parameter in RRV framework.

2.1. Realized range-based volatility measure

Using the extreme value method of intra-day return by the high-low range, Martin and Dick (2007) provide a more efficient measure called the realized range-based volatility (RRV), which uses the high-low range instead of the squared return. In their study, RRV significantly improves over RV due to the price range of intraday carrying more information than the closing prices. In order to deal with microstructure issues, we can use a bias-correction procedure to account for the effects of microstructure frictions based upon scaling the realized range with the average level of the daily range. The RRV has been proved to have a lower mean-squared error than RV by simulation experiments (Martin and Dick, 2007). Their result shows RRV is robust to microstructure noise.

Initially, assuming that the oil futures price, p_t, follows a geometric Brownian motion, considering the equidistant partition $0=t_0 < t_1<\ldots<t_M=1$, where $t_i=i/M$,$t_{i-1}\leq t$, $s\leq t_i$, and $\Delta=1/M$, the intraday range at sampling times t_{i-1} and $t_i (i = 1,2, \ldots, M)$ is

$$s_{p,\Delta i}=\sup\{p_t - p_s\}. \tag{1}$$

The estimator RRV for the interval [0, 1] can be expressed as:

$$RRV_t^{\Delta} = \frac{1}{\lambda_{2,m}} \sum_{j=1}^{1/\Delta} s^2_{(t-1)+j*\Delta,\Delta}, \tag{2}$$

where $\lambda_{r,m} = \mathbf{E}(s^r_{w,m}).\lambda_{r,m}$ is the r_{th} moment of the range of Brownian motion over a unit interval. When $\triangle\to 0$, realized range-based volatility (RRV) and the realized range-based bi-power variation (RBV) are specified as below (Christensen and Podolskij, 2006; Christensen and Podolskij, 2007):

$$RRV^{\Delta}_{t,m} \to \int_0^t \sigma^2(s)ds + \frac{1}{\lambda_{2,m}} \sum_{0<s\leq t} k^2(s). \tag{3}$$

$$RBV_{t,m} = \frac{1}{\lambda^2_{1,m}} \sum_{j=2}^{1/\Delta} |s_{(t-1)+j*\Delta}||s_{(t-1)+(j-1)*\Delta}| \to \int_0^t \sigma^2(s)ds. \tag{4}$$

From the Eqs. (3) and (4), we have,

$$RRV_{t,m} = \lambda_{2,m}RRV^{\Delta}_{t,m} + (1-\lambda_{2,m})RBV_{t,m} \to \int_0^t \sigma^2(s)ds + \sum_{0<s\leq t} k^2(s), \tag{5}$$

where $\int_0^t \sigma^2(s)ds$ represents continuous path component and

$\sum_{0<s\le t} k^2(s)$ represents the jump component. Therefore, to ensure that estimates of all daily jump variation are non-negative, the jump component J_t is given by,

$$J_t = \max(RRV_{t,m} - RBV_{t,m}, 0). \tag{6}$$

Furthermore, the "significant" jump can be tested using this modified ratio-statistic,

$$Z_{t,m} = \frac{\sqrt{n}\left(1-\frac{RBV_{t,m}}{RRV_{t,m}}\right)}{\sqrt{v_m \max\left\{\frac{1}{t}, \frac{RQQ_{t,m}}{(RBV_{t,m})^2}\right\}}} \to N(0, 1), \tag{7}$$

where m=1/Δ, t is the length of the sample period, and v_m is determined from $\lambda_{1,m}$, $\lambda_{2,m}$, $\lambda_{3,m}$, and $\lambda_{4,m}$, and that are computed by a Brownian motion with 1,000,000 replications (Christensen and Podolskij, 2006; Christensen and Podolskij, 2012). Therefore, the "significant" jumps are identified by the realizations of $Z_{t,m}$ in excess of some critical value, say Φ_α,

$$CJ_{t,m} \equiv I(Z_{t,m} > \Phi_\alpha)\cdot[RRV_{t,m} - RBV_{t,m}], \tag{8}$$

where $I(\cdot)$ is an indication function.

2.2. Modelling realized range-based volatility

Based on realized range-based volatility (e.g., Andersen et al., 2007; Corsi, 2009), we have the specification of HAR-RRV,

$$RRV_{t+1} = c + \beta_d RRV_t + \beta_w RRVW_t + \beta_m RRVM_t + \varepsilon_{t+1}, \tag{9}$$

where c is the constant and ε_{t+1} refers to error term. HAR-RRV has three explanatory variables: lagged daily realized range-based volatility (RRV_t), lagged weekly realized range-based volatility ($RRVW_t$) and lagged monthly realized range-based volatility ($RRVM_t$). The weekly indicator $RRVW_t$ is the average daily RRV from day$t-4$ to day t. The monthly indicator $RRVM_t$ is the average daily RRV from day $t-21$ to day t.

Tseng et al. (2012) states that overnight return improves forecasting RRV. Accordingly, we use HAR-RRV-ONI model with overnight return indicating overnight information to forecast RRV. Based on Eq. (9), by considering overnight information, HAR-RRV-ONI model is constructed by adding overnight return, ONI_t, as a new explanatory variable. HAR-RRV-ONI model can be expressed as:

$$RRV_{t+1} = c + \beta_d RRV_t + \beta_w RRVW_t + \beta_m RRVM_t + \beta_{ONI} ONI_t + \varepsilon_{t+1}, \tag{10}$$

where the overnight return ONI_t is the return between the closing price and opening price.

HAR-RRV-J model is extended by adding the jump component, J_t, to HAR-RRV model and thus we have the specification of HAR-RRV-J as follows:

$$RRV_{t+1} = c + \beta_d RRV_t + \beta_w RRVW_t + \beta_m RRVM_t + \beta_j J_t + \varepsilon_{t+1}, \tag{11}$$

where the jump component is referred to $J_t = max\,(RRV_t - RBV_t, 0)$, $RBV_t = \mu_1^{-2}\sum_{j=2}^{M}|r_{t,j-1}||r_{t,j}|$, and $\mu_1 = \sqrt{2/\pi} = E(|Z|)$. Based on Andersen et al. (2007), we can extend HAR-RRV model to HAR-RRV-CJ model incorporating continuous sample path and jump components, and it can be written as:

$$RRV_{t+1} = c + \beta_d CRRV_{t,\alpha} + \beta_w CRRVW_{t,\alpha} + \beta_m CRRVM_{t,\alpha} + \beta_j CJ_{t,\alpha} + \beta_{jw} CJW_{t,\alpha} + \beta_{jm} CJM_{t,\alpha} + \varepsilon_{t+1}, \tag{12}$$

where $CRRV_{t,\alpha}$, $CRRVW_{t,\alpha}$ and $CRRVM_{t,\alpha}$ are the daily, weekly, and monthly averages of continuous sample path, respectively, and similarly $CJ_{t,\alpha}$,$CJW_{t,\alpha}$, $CJM_{t,\alpha}$, refer to the daily, weekly, and monthly jump components.

Let $I(\cdot)$ be the indicator function, and thus these components can be calculated as follows:

$$CRRV_{t,\alpha} = I(Z_t \le \Phi_\alpha)\cdot RRV_t + I(Z_t > \Phi_\alpha)\cdot BPV_t, \tag{13}$$

$$CRRVW_{t,\alpha} = I(Z_t \le \Phi_\alpha)\cdot RRVW_t + I(Z_t > \Phi_\alpha)\cdot BPVW_t, \tag{14}$$

$$CRRVM_{t,\alpha} = I(Z_t \le \Phi_\alpha)\cdot RRVM_t + I(Z_t > \Phi_\alpha)\cdot BPVM_t, \tag{15}$$

$$CJ_{t,\alpha} = I(Z_t > \Phi_\alpha)\cdot(RRV_t - BPV_t). \tag{16}$$

Accordingly, the weekly jump component is calculated as $CJW_{t,\alpha} = I(Z_t > \Phi_\alpha)\cdot[RRVW_t - BPVW_t]$, and the monthly jump component is calculated as $CJM_{t,\alpha} = I(Z_t > \Phi_\alpha)\cdot[RRVM_t - BPVM_t]$.

Besides aforementioned models, by considering leverage terms related to negative returns (McAleer and Medeiros, 2008), we can construct the model LHAR-RRV-CJ and specify it as:

$$RRV_{t+1}^{(d)} = c + \beta_d RRV_t + \beta_w RRVW_t + \beta_m RRVM_t + \gamma_d r_{d,t}^- + \gamma_w r_{w,t}^- + \gamma_m r_{m,t}^- + \varepsilon_{t+1}, \tag{17}$$

where $r_{d,t}^- = r_t I(r_t<0)$ is the negative part of the daily log return,$I(\cdot)$ is the indicator function, and

$$r_{w,t}^- = \frac{1}{5}(r_t+\ldots+r_{t-4})I(r_t+\ldots+r_{t-4}<0), \tag{18}$$

$$r_{m,t}^- = \frac{1}{22}(r_t+\ldots+r_{t-21})I(r_t+\ldots+r_{t-21}<0). \tag{19}$$

In our empirical study, to forecast oil futures volatility, aforementioned five individual models, such as HAR-RRV, HAR-RRV-ONI, HAR-RRV-J, HAR-RRV-CJ, and LHAR-RRV-CJ, are employed.

2.3. Combined models

Although some individual models have relatively good predictability, their forecasting abilities are changeable over time. Hence, combining several models into a single model is effective in forecasting. There are several classes of combined models: the mean combination, discounted MSFE forecasts, shrinkage forecasts, and 'most recently best' methods (Stock and Watson, 2004).

The mean combination takes the equal-weighted average of the forecasts generated by different individual models. It has two extensions: the median and the trimmed mean. The median combination uses the median of volatility forecasts while the trimmed mean is calculated with 5% symmetric trimming. For example, the simple mean combination equation can be written as:

$$\hat{\sigma}_{t+1,\text{combine}}^2 = \frac{1}{5}[\hat{\sigma}_{t+1,\text{HAR-RRV}}^2 + \hat{\sigma}_{t+1,\text{HAR-RRV-J}}^2 + \hat{\sigma}_{t+1,\text{HAR-RRV-CJ}}^2 + \hat{\sigma}_{t+1,\text{HAR-RRV-ONI}}^2 + \hat{\sigma}_{t+1,\text{LHAR-RRV-CJ}}^2], \tag{20}$$

where $\hat{\sigma}_{t+1,\text{combine}}^2$ denotes 1-step ahead combination forecast at time t +1, and the five terms within the brackets represent the forecasts from the models HAR-RRV, HAR-RRV-J, HAR-RRV-CJ, HAR-RRV-ONI, and LHAR-RRV-CJ, respectively.

The shrinkage forecasts have two methods to obtain the weights. One is averaging the weights of OLS estimator and the other is equal weighting. Following Stock and Watson (2004), the shrinkage forecasts can be written as:

$$\omega_{it} = \lambda\hat{\beta}_{it} + (1-\lambda)(1/n), \tag{21}$$

where $\hat{\beta}_{it}$ represents the ith coefficient estimated from OLS regression of forecasts series generated from individual models. $\lambda = max\,\{0, 1 - \kappa[n/(t-h-T_0-n)]\}$, where n is the number of individual forecasts, T_0 is the starting period which individual out-of-sample forecasts are computed from. h represents the number of steps ahead forecasting. κ is a given constant which is used to control the amount of shrinkage towards equal weighting. The shrinkage forecasts are evaluated for κ = 0.25, 0.5, and 1. The weight κ could be estimated using empirical Bayes methods and larger value of κ indicates faster shrinkage.

In the empirical study, we apply three mean combination forecasts (equal-weighted average, the median, and the trimmed mean) that are denoted as M1, M2 and M3, and three shrinkage forecasts given κ = 0.25, 0.5, 1 that are denoted M4, M5 and M6, respectively.

2.4. Dynamic model averaging approach

The individual HAR-RRV-type models are presented in Section 2.2 and combined models presented in Section 2.3 are regression models with constant weights. However, in a volatile market, asset prices are unstable over time, so that the most appropriate model at different points is not easy to be determined. This problem can be overcome by Bayesian model averaging (BMA) and Bayesian model selection (BMS) procedures (Cremers, 2002; Koop and Potter, 2004; Wright, 2008). Raftery et al. (2012) argue that BMA sets fixed weight for each model and is still restricted to static problems. BMS selects the model by maximizing the likelihood probability of forecasting at the next period. Although the combined models with constant weights can perform better than the individual models, they still cannot capture the changing weights of individual models over time. Thus, how time-varying combined models is applied in RRV framework has to be addressed. Raftery et al. (2010) propose dynamic model averaging (DMA) and dynamic model selection (DMS) approach, which allows the models and their coefficients to change over time. Due to the flexibility of DMA, this study first considers incorporating DMA into RRV framework to combine the forecasting models. The discussion of DMA approach can be found in Appendix A.

3. Data

Since 5-min high-frequency data has been proved to be a rule-of-thumb and it is the best trade-off between market microstructure noise and measurement accuracy (e.g., Andersen and Bollerslev, 1998; Corsi et al., 2010; Sévi, 2014; Liu et al., 2015), in this study, we use 5-min high-frequency data of the Light Sweet Crude Oil (WTI) futures contract with a maturity of one month from the NYMEX. Since the most traded commodity futures and most of oil-based derivatives throughout the world are priced with respect to the WTI contract. Our data are collected from the Tick DATA, covering the period from January 2, 2007 to May 9, 2014. Totally 1851 daily observations are obtained after removing days with a shortened trading session or too few transactions[1]. For robustness, we also examine stock market by using representative the S & P 500 index[2].

Our sample data of oil futures is divided into two subgroups: a) in-sample data used to estimate the models, covering the first 1000 trading days (from January 2, 2007 to December16, 2010); and b) out-of-sample data used for model evaluation, covering the rest 850 trading days (from December17, 2010 to May 9, 2014). The estimation period is then rolled forward by dropping the most distant day and adding the new day. Thus, the in-sample data remains at a fixed length and the forecasts do not overlap[3]. Besides, our sample data of the S & P 500 index is from January 4, 2000 to April 29, 2016. This data is for the trading time of each business day between 9:30:00 and 16:00:00. After removing days with shortened trading sessions or too few transactions, high-frequency data for 4012 business days are obtained, and the last 992 days are used for the model evaluation.

Table 1 presents the descriptive statistics of several volatility series based on oil futures price and the S & P 500 index. All realized measures and jumps variations series are significantly skewed and leptokurtic at the 1% level, suggesting that their distributions have fat-tail and are more peaked than Gaussian distribution. The Jarque-Bera statistic (Jarque and Bera, 1987) further demonstrates that the null hypothesis of normality is rejected at the 1% significance level. The above results indicate that all of series we have discussed are not normally distributed. The Ljung-Box statistic Q(5) (Ljung and Box, 1978) for serial correlation shows that the null hypotheses of no autocorrelation up to the 5th order are rejected for most of series, which suggest that there exist series correlation in the realized measures and jumps variations series.

4. Empirical results

This section first introduces the forecasting methodology of out-of-sample rolling method and the model confidence set (MCS) test, and then presents in-sample forecasts of individual HAR-RRV-type models and out-of-sample forecasts, which include the individual HAR-RRV-type models, combined models with constant weights, BMA/BMS, and DMA/DMS.

4.1. Forecasting methodology and forecasts evaluation

In this section, we briefly introduce the forecasting methodology of out-of-sample rolling method and forecasts evaluation methods. Since the forecasting models are proven to be sensitive the error statistic used to assess the accuracy of the forecasts. For robustness, the models are assessed not only by error statistics but also assessed by mean mixed statistics (MME) and the model confidence set (MCS) proposed by Hansen et al. (2011).

Forecasting models are usually evaluated by various loss functions. However, these loss functions cannot provide any information on whether the differences of forecasting losses among models are significant. Hansen (2005) proposes a superior predictive ability (SPA) test and solves the problem of comparing multiple forecasts. SPA method is better than the aforementioned loss functions to evaluate the forecasts, but it requires providing a benchmark model. Subsequently, Hansen et al. (2011) propose a much more advanced method, called advanced model confidence set (MCS) test. The MCS test has several attractive advantages over both conventional tests and SPA method. First, the MCS test does not need to specify a benchmark which is very useful in applications without an obvious benchmark. Secondly, the MCS procedure allows for the possibility of more than one "best" model. If p-values of both statistic tests are larger than 0.1, the corresponding model is called "surviving" model. It means that those "surviving" models have better performance than the models removed by MCS test (Hansen et al., 2011). Because of those advantages, MCS test is widely used to compare forecasting performance, and hence we will apply MCS as the criteria for model comparison. In addition, the specification of mean mixed statistics (MME) and the results of error statistics are provided in Appendix B.

To quantitatively evaluate the forecasting accuracy of the models, some of the popular loss functions can be calculated as:

$$HMSE = M^{-1}\sum_{t=1}^{M}(1-\hat{\sigma}_t^2/RRV_t)^2, \tag{22}$$

$$HMAE = M^{-1}\sum_{t=1}^{M}|1-\hat{\sigma}_t^2/RRV_t|, \tag{23}$$

where $\hat{\sigma}_t^2$ refers to the volatility forecasts generated by different individual models and their combinations and M denotes the number of forecasting data points (M = 850). RRV_t is the proxy for actual market volatility in out-of-sample period.

The MCS method has obvious advantages mentioned over conven-

[1] The trading hours of crude oil futures are from 6:00 PM until 5:15 PM, Sunday through Friday with a 45-minute break each day between 5:15 PM (current trade date) and 6:00 PM (next trade date), New York Time.

[2] The empirical results of 10- and 15- min high-frequency data can also be provided according to the requests.

[3] More details and discussions about the optimization of estimation window can be seen in Rossi and Inoue (2012). In a robustness check, we have tested estimation window for 900, 1000, and 1100 days. The results of these different windows are consistent.

Table 1
Descriptive statistics of each volatility series.

Data	Variables	Mean	St.dev.	Skewness	Kurtosis	Jarque-Bera	Q(5)
Crude oil futures	*RRV*	3.005	4.262	3.941***	19.335***	33225.704***	5453.734***
	RRVW	2.892	4.155	3.940***	19.034***	32340.317***	5428.804***
	RRVM	2.772	3.921	3.842***	18.006***	29062.655***	5428.804***
	CRRV	2.726	3.906	3.988***	20.329***	36164.290***	5408.288***
	J	0.324	0.775	7.285***	73.919***	432583.731***	455.498***
	CJ	0.113	0.579	12.325***	198.908***	3046388.890***	15.129***
The S & P 500 index	*RRV*	0.808	1.784	13.319***	319.558***	17193465.119***	7165.563***
	RRVW	0.807	1.488	7.548***	79.897***	1105496.111***	15744.426***
	RRVM	0.805	1.317	5.978***	5.978***	45.382***	19439.385***
	CRRV	0.622	1.560	15.078***	384.999***	24936431.868***	7199.179***
	J	0.187	0.394	7.191***	83.212***	1192371.749***	1868.838***
	CJ	0.187	0.394	7.188***	83.152***	1190689.646***	1848.769***

Notes: The null hypothesis are “Skewness = 0” and “Kurtosis= 3”. Kurtosis is the excess kurtosis. The Jarque-Bera statistic tests the null hypothesis of normality for the distribution of the series. $Q(n)$ is the Ljung-Box statistic for up to 5th order serial correlation.
*** denotes rejections of null hypothesis at the significance level of 1%. All of the realized measures and jumps variations series of Table 1 are multiply by 10000.

tional tests such as the superior predictability and the “Reality check” (Hansen and Lunde, 2005; White, 2000). MCS test is used to select a subset of models with significant predictability from a set of models. In our empirical study, we use two popular statistic tests of MCS: the range statistic (T_R) and semi-quadratic statistic (T_{SQ}). If p-values of both of statistic tests are larger than γ, the corresponding models are “surviving” models, which means that those “surviving” models achieve better performance than models which have been removed by MCS test in forecasting.

4.2. In-sample analysis

Table 2 exhibits the in-sample estimates of the HAR-RRV-type models using oil futures and the S & P 500 index. For oil futures, the estimation parameters of weekly and monthly RRV of individual models are significant at the 5% level. For the S & P 500 index, all the parameters of daily, weekly and monthly RRV except one are significant at the 5% level. That suggests a strong persistence in RRV dynamics. In addition, the coefficient of the weekly parameter is larger than that of the monthly and daily parameters in each model based on

Table 2
In-sample parameter estimates of individual HAR-RRV-type models (α=0.05).

Data	Coefficient	HAR-RRV	HAR-RRV-J	HAR-RRV-CJ	LHAR-RRV-CJ	HAR-RRV-ONI
Oil futures price	c	0.0000	0.0000	0.0000	0.0000**	0.0000
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
	β_d	0.0232	0.1189**	0.1062**	-0.0072	0.0121
		(0.0278)	(0.0349)	(0.0365)	(0.0273)	(0.0279)
	β_w	0.6771**	0.6525**	0.7007**	0.5863**	0.6927**
		(0.0522)	(0.0522)	(0.0659)	(0.0538)	(0.0522)
	β_m	0.2730**	0.2592**	0.1128	0.2650**	0.2696**
		(0.0440)	(0.0438)	(0.0621)	(0.0549)	(0.0439)
	$\beta_j/\gamma_d/\beta_{ONI}$		-0.3517**	-0.2106**	0.3870	-0.0008**
			(0.0778)	(0.0707)	(0.2491)	(0.0002)
	β_{jw}/γ_w			0.3581**	-0.0020**	
				(0.1643)	(0.0003)	
	β_{jm}/γ_m			1.0960**	-0.0021**	
				(0.2547)	(0.0007)	
	F-test	2178.9611**	1656.5569**	1110.9901**	1165.4377**	1646.5159**
	R^2	0.7798	0.7822	0.7834	0.7812	0.7914
The S & P 500 index	c	0.0000**	0.0000**	0.0000**	0.0000**	0.0000**
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
	β_d	0.1966**	0.2722**	0.2079**	0.0908**	0.1967**
		(0.0191)	(0.0211)	(0.0228)	(0.0186)	(0.0188)
	β_w	0.5144**	0.5131**	0.7275**	0.4040**	0.5185**
		(0.0323)	(0.0321)	(0.0394)	(0.0307)	(0.0318)
	β_m	0.1950**	0.1932**	-0.0473	0.1716**	0.1919**
		(0.0286)	(0.0284)	(0.0385)	(0.0333)	(0.0281)
	$\beta_j/\gamma_d/\beta_{ONI}$		-0.5156**	-0.0890	-0.1665	-0.0095**
			(0.0642)	(0.0672)	(0.1499)	(0.0008)
	β_{jw}/γ_w			-0.8172**	0.0001	
				(0.1469)	(0.0003)	
	β_{jm}/γ_m			1.9281**	-0.0134**	
				(0.1849)	(0.0007)	
	F-test	1486.8002	1148.9279	804.9915	932.4555	1186.5361
	R^2	0.5266	0.5342	0.5466	0.5827	0.5422

Notes: This table reports the estimates of Eqs. (17), (18), (19), (20), and (25). The numbers in parentheses are standard errors of the estimations.
** denotes rejections of null hypothesis at the 5% significance level.

two data, implying that the mid-term investor has a greater impact on future volatility than short- and long-term investors. Moreover, most of the jump components in these models are negative and show a significantly negative impact on future realized range-based volatility. The fourth model LHAR-RRV-CJ combines heterogeneity into realized ranged-based volatility, leverage, and jumps. From its estimation, we can find that the coefficients of negative returns are significant and there exists the leverage effect. This emphasizes that the leverage effect is important for modelling and forecasting volatility.

The fifth individual model HAR-RRV-ONI is constructed by adding the overnight information as an explanatory variable. The evidence suggests that the overnight information has a significantly negative impact on future volatility both for oil futures and the S & P 500 index. Finally, the R-squared (R^2) values for oil futures are above 77% while the R-squared (R^2) values for the S & P 500 index are above 52%, indicating good fitness of these models to the volatility and better fitness to the oil futures. Meanwhile RRV models with jumps and overnight information obtain higher R^2, implying better fitness of data than that of the simple HAR-RRV model.

4.3. Out-of-sample evaluation

As stated in the literature, in-sample predictability is not constant and changes over time while the out-of-sample performance of a model has a greater future forecasting ability. Thus, out-of-sample predictability is more accepted by both researchers and market participants. The forecasting performance of individual models, individual models and combined models M1-M6, individual models and BMA/DMA, and all the models used in our empirical study including DMA approach, are compared in group by MCS test in Tables 3–6.

Table 3 presents the results of our five basic HAR-RRV-type models, which we have discussed in Section 2.2. Under the *HMAE* and *HMAE* loss functions, we find that HAR-RRV-J and LHAR-RRV-CJ are better than others for forecasting oil futures. HAR-RRV-CJ and HAR-RRV-ONI are significantly better for forecasting the S & P 500 index. The empirical results indicate that the best forecasting model depends on the financial market and the jump component can improve the model's forecasting accuracy.

However, there are some issues using individual models for forecasting. First, individual models may be differently affected by structural breaks over time. Second, they may also be subject to misspecification bias of unknown form. Thirdly, underlying forecasts may be based on different loss functions. By contrast, combining forecasts with different degrees of adaptability can reduce the above issues and outperform forecasts from individual models (Hendry and Clements, 2004; Stock and Watson, 2004). In our empirical study, six combined models with constant weights, labeled as M1-M6, are employed. M1, M2, and M3 represent three simple combined models (the mean, the median, and the trimmed mean) and M4, M5, and M6 denote three shrinkage forecasts given κ = 0.25, 0.5, and 1, respectively.

Table 4 compares the five individual models to six combined models with constant weights under *HMSE* and *HMAE*. For oil futures price, LHAR-RRV-CJ and the three combined models of M1, M3, and M6 are significantly better than others. And hence the models M1, M3, and M6 are successful combinations. For the S & P 500 index, none of the individual models are significant while the combined models, M1 and M3 outperform others. M1, which refers to the combined model using equal-weighted average, shows the strong predictability across markets. Also, this evidence is consistent with Graefe et al. (2014), which states that simple average is expected to perform better. Overall, the combined models perform better than individual models. The combined model using equal-weighted average is robust across markets and recommended, because it not only outperforms but also it is easy to describe, understand, and implement.

Table 5 presents the results for comparing the five individual models with four specific models, namely BMA, BMS, DMA and DMS, which can describe the models' time-varying features. In this study, DMA is first applied as the combined model with time-varying weights to combine RRV forecasts generated from different models. The time-varying weights for different individual HAR-RRV-type models are computed by the process introduced in Section 2.4. There is strong evidence that DMA approach (combined model with time-varying weights) outperforms individual models in Table 5 both for oil futures and the S & P 500 index. As there are rapid changes for several times in our sample period, DMA method can be flexible to those changes. In addition, DMA approach shows its advantage in capturing the dynamic of models and parameters. Hence, using DMA as a combined model can help generate more accurate forecasts than individual models.

Furthermore, given the above empirical results, we want to find the best method among all the models. In this part, we evaluate the forecasting performance of all models including individual HAR-RRV-type models, combined models with constant weights, and other combined models with time-varying weights by using the MCS test. The empirical results of comparing all the models are reported in Table 6. For oil futures market, DMA approach significantly outperforms both individual models and combined models, including M1 which is always used as the benchmark model in combined models. For the S & P 500 index, four combined models DMA, DMS, M1 and M3 perform better than others. Among these four models, all the p-values of DMA are significant under the two statistics. Therefore, combined models show better forecast ability than individual models and DMA approach is the best one among combined models.

For robustness check, we test DMA by using alternative forgetting factor values with λ = 0.99, α = 0.94 in Appendix B Table B1. We test all the models based on 10-min and 15-min high-frequency data and present the overall comparison in Appendix B Tables B2 and B3. In addition, we evaluate the models under different criteria such as error statistics and mean mixed statistics (MME) that are shown in Appendix

Table 3
MCS results for comparing individual HAR-RRV-type models.

Model	Crude future oil				the S & P500			
	HMSE		*HMAE*		*HMSE*		*HMAE*	
	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}
HAR-RRV	0.0000	0.0000	0.0000	0.0000	0.1745	0.1118	0.0173	0.0165
HAR-RRV-J	**0.3638**	**0.3638**	**0.3663**	**0.3663**	0.0000	0.0000	0.0000	0.0000
HAR-RRV-CJ	0.0000	0.0000	0.0000	0.0000	**1.0000**	**1.0000**	**1.0000**	**1.0000**
LHAR-RRV-CJ	**1.0000**	**1.0000**	**1.0000**	**1.0000**	0.0003	0.0000	0.0000	0.0000
HAR-RRV-ONI	0.0000	0.0000	0.0133	0.0021	**0.3911**	**0.3911**	**0.3324**	**0.3324**

Notes: The MCS *p*-values are computed according to the statistics $T_R$and T_{SQ}. The numbers in this table are the MCS *p*-values for the different forecasts. The larger the number is, the better the corresponding model performs. The p-values, which are larger than 0.25 indicated in bold and under and underline, denote that the corresponding models are the best based on MCS test.

Table 4
MCS results for comparing HAR-RRV-type individual models with combined models.

Model	Crude future oil				The S & P 500 index			
	HMSE		*HMAE*		*HMSE*		*HMAE*	
	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}
HAR-RRV	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
HAR-RRV-J	**0.2420**	0.2560	**0.4571**	0.2420	0.0000	0.0000	0.0000	0.0000
HAR-RRV-CJ	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LHAR-RRV-CJ	**1.0000**	**1.0000**	**1.0000**	**1.0000**	0.0000	0.0000	0.0000	0.0000
HAR-RRV-ONI	0.0000	0.0000	**0.4571**	0.0000	0.0000	0.0000	0.0000	0.0000
M1	**0.5932**	**0.6527**	**0.5022**	**0.5932**	**0.5429**	**0.5429**	**1.0000**	**1.0000**
M2	0.0002	0.0030	**0.4571**	0.0002	0.0000	0.0000	0.0000	0.0000
M3	**0.5932**	**0.6527**	**0.5022**	**0.5932**	**1.0000**	**1.0000**	0.0933	0.0933
M4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
M5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
M6	**0.3347**	**0.4124**	**0.5022**	**0.3347**	0.0000	0.0000	0.0000	0.0000

Notes: The MCS *p*-values are computed according to the statistics $T_R$and T_{SQ}. The numbers in this table are the MCS p-values for the different forecasts. The larger the number is, the better the corresponding model performs. The p-values, which are larger than 0.25 indicated in bold and under and underline, denote that the corresponding models are the best based on MCS test. M1-M6 represent the combined models: equal-weighted average, the median combination, the trimmed mean, and the shrinkage forecasts for κ = 0.25, 0.5, and 1.

Table 5
MCS results for comparing HAR-RRV-type individual models with BMA/DMA.

Model	Crude future oil				the S & P500			
	HMSE		*HMAE*		*HMSE*		*HMAE*	
	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}
BMA	0.0000	0.0000	0.0000	0.0000	0.1020	0.0000	0.0299	0.0000
BMS	0.0000	0.0000	0.0000	0.0000	0.1020	0.0019	0.0299	0.0141
DMA	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
DMS	0.1695	0.1224	0.1237	0.0624	0.1020	0.0260	0.0299	0.0253
HAR-RRV	0.0090	0.0007	0.0000	0.0000	0.1020	0.0008	0.0299	0.0000
HAR-RRV-J	0.1423	0.0878	0.0085	0.0051	0.1020	0.0000	0.0299	0.0000
HAR-RRV-CJ	0.0090	0.0000	0.0000	0.0000	0.1020	0.0019	0.0299	0.0000
LHAR-RRV-CJ	0.1695	0.1224	0.1237	0.0624	0.1020	0.0000	0.0299	0.0000
HAR-RRV-ONI	0.0100	0.0061	0.0000	0.0001	0.1020	0.0019	0.0299	0.0000

Notes: The MCS *p*-values are computed according to the statistics $T_R$and T_{SQ}. The numbers in this table are the MCS p-values for the different forecasts. The larger the number is, the better the corresponding model performs. The numbers larger than 0.25 which are indicated in bold and underline demonstrates that the corresponding models are the best models in MCS test. M1-M6 represent the combined models: equal-weighted average, the median combination, the trimmed mean, and the shrinkage forecasts for κ = 0.25, 0.5, and 1.

Table 6
MCS results for comparing all the models.

Model	Crude oil futures				The S & P500 index			
	HMSE		*HMAE*		*HMSE*		*HMAE*	
	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}
BMA	0.000	0.000	0.000	0.000	0.088	0.000	0.053	0.062
BMS	0.000	0.000	0.000	0.000	0.088	0.000	0.053	0.062
DMA	**1.000**	**1.000**	**1.000**	**1.000**	**0.584**	**0.789**	**1.000**	**1.000**
DMS	0.216	0.161	0.124	0.062	**0.576**	**0.317**	**0.405**	0.216
HAR-RRV	0.115	0.000	0.008	0.000	0.088	0.000	0.053	0.000
HAR-RRV-J	0.115	0.114	0.009	0.008	0.088	0.000	0.053	0.000
HAR-RRV-CJ	0.115	0.000	0.008	0.000	0.088	0.000	0.053	0.000
LHAR-RRV-CJ	0.216	0.161	0.124	0.062	0.088	0.000	0.053	0.000
HAR-RRV-ONI	0.115	0.000	0.008	0.000	0.088	0.000	0.053	0.000
M1	0.216	0.161	0.008	0.008	**0.584**	**0.789**	**0.405**	0.216
M2	0.115	0.004	0.008	0.000	0.088	0.000	0.053	0.000
M3	0.115	0.161	0.008	0.008	**1.000**	**1.000**	0.405	0.216
M4	0.000	0.000	0.000	0.000	**0.576**	0.000	0.053	0.000
M5	0.000	0.000	0.000	0.000	**0.576**	0.000	0.053	0.000
M6	0.115	0.139	0.008	0.008	**0.576**	0.000	0.053	0.000
GARCH(1,1)	0.000	0.000	0.000	0.000	0.088	0.000	0.053	0.000
EGARCH	0.000	0.000	0.000	0.000	0.088	0.000	0.053	0.000

Notes: The MCS *p*-values are computed according to the statistics $T_R$and T_{SQ}. The numbers in this table are the MCS p-values for the different forecasts. The larger the number is, the better the corresponding model performs. The p-values, which are larger than 0.25 indicated in bold and under and underline, denote that the corresponding models are the best based on MCS test. M1-M6 represent the combined models: equal-weighted average, the median combination, the trimmed mean, and the shrinkage forecasts for κ = 0.25, 0.5, and 1.

B Tables B4-B7. DMA is compared with alternative one by DM statistics in Table B8. Finally, we investigate the forecasting performance of DMA at several estimation window sizes and the results are exhibited in Appendix B Tables B9 and B10. The above results are consistent with the MCS test. Therefore, we conclude that the predictability of DMA is robust to high-frequency data for forecasting RRV of financial assets.

In summary, we can gain clear and strong conclusions based on oil futures price and the S & P 500 index. First, combined models with constant weights could beat individual models, which mean that the combined methods can gain a higher prediction accuracy. Especially, combined model with simple average is recommended. Second, the DMA approach, used as the combined model with time-varying weights, can not only generate more accurate forecasts, but also beat individual models and combined models with constant weights. Thus, DMA is successfully used as a combined method in RRV framework. Applying DMA into RRV framework can significantly improve the forecasting accuracy.

5. Conclusions

Since the realized range-based volatility (RRV) is proved to be a better measurement than the realized volatility (RV), this study uses RRV as the proxy for oil futures volatility. To investigate the time-varying property of HAR-RRV models' parameters, this paper applies DMA approach as a combined model with time-varying weights into RRV framework.

At first, this study constructs five individual HAR-RRV-type models including HAR-RRV-ONI considering overnight information. Moreover, based on the individual models, combined models with constant weights and time-varying weights are constructed. Finally, the performance of three types of the models is compared by various methods including MCS test. Our findings demonstrate that the HAR-RRV-type models can successfully capture the long-term memory behavior of volatility in oil futures market. Combined models systematically perform better than individual models. In particular, using DMA to combine the forecasts of HAR-RRV models can significantly improve the forecasting accuracy. DMA approach can beat both individual models and combined models with constant weights, including the combined model with equal-weighted average, which is usually used as the benchmark. The results highlight the significance of using DMA as a combined model with time-varying weights. By allowing both the models and their coefficients changing over time, DMA approach shows several benefits to be incorporated into RRV framework and opens a new path for forecasting technique in real time.

Acknowledgments

The authors are grateful to the Co-Editor and the anonymous referees for providing thoughtful comments and suggestions that enhanced the quality of this paper. The authors are also grateful to the financial support from the China Scholarship Council [2015070000042]; Natural Sciences and Engineering research Council of Canada; the Natural Science Foundation of China [71371157, 71671145]; the humanities and social science fund of ministry of education [14YJC790073, 15YJA790031 and 16YJA790062]; the young scholar fund of science & technology department of Sichuan province [2015JQO010]; and Fundamental research funds for the central universities [26816WCX02].

Appendix A

Since DMA approach is the main model applied, more information is given in this section. The dynamic model averaging (DMA) and dynamic model selection (DMS) approach, proposed by Raftery et al. (2010), allows both the factors and their coefficients changing over time.

Following Raftery et al. (2010), DMA can be specified as:

$$RRV_{t-1} = \beta_{t-1}^{(k)} z_{t-1}^{(k)} + \varepsilon_{t-1}^{(k)}, \tag{A1a}$$

$$\beta_t^{(k)} = \beta_{t-1}^{(k)} + \eta_{t-1}^{(k)}, \tag{A1b}$$

where $k = 1, \ldots, K$, the errors $\varepsilon_{t-1}^{(k)} \sim N(0, H_{t-1}^{(k)})$ and $\eta_{t-1}^{(k)} \sim N(0, Q_{t-1}^{(k)})$ are mutually independent at all of the leads and lags, the vector $z_{t-1}^{(k)}$ is the set of predictors of model k including the intercept, and $\beta_t^{(k)}$ refers to the time-varying coefficients of forecasts from individual models.

In a single case, for given values of H_t and Q_t, standard filtering results can be used to compute recursive estimation or forecasting. Kalman filtering begins with the result:

$$\beta_{t-1} \mid RRV^{t-1} \sim N\left(\hat{\beta}_{t-1}, \sum_{t-1|t-1} \cdot\right). \tag{A2}$$

Kalman filtering proceeds using:

$$\beta_t \mid RRV^{t-1} \sim N\left(\hat{\beta}_{t-1}, \sum_{t|t-1} \cdot\right). \tag{A3}$$

where

$$\sum_{t|t-1} \cdot = \sum_{t-1|t-1} \cdot + Q_{t.} \tag{A4}$$

Involving the forgetting factor λ $(0 \le \lambda \le 1)$ that refers to a gradual evolution of coefficients, Eq. (A4) can be replaced by:

$$\sum_{t|t-1} \cdot = (1/\lambda) \sum_{t-1|t-1} \cdot. \tag{A5}$$

Thus, we have $Q_t = (1 - \frac{1}{\lambda}) \sum_{t-1|t-1} \cdot$. Estimation can be completed by the updating equation:

$$\beta_t \mid RRV^t \sim N\left(\hat{\beta}_t, \sum_{t|t} \cdot\right), \tag{A6}$$

where

$$\hat{\beta}_t = \hat{\beta}_{t-1} + \Sigma_{t|t-1}(H_t + z_t z_t \Sigma_{t|t-1} z_t')^{-1}(RRV_t - z_t \hat{\theta}_{t-1}), \tag{A7}$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1} z_t (H_t + z_t \Sigma_{t|t-1} z_t')^{-1} z_t \Sigma_{t|t-1}. \tag{A8}$$

By using the predictive distribution,

$$RRV_t \mid RRV^{t-1} \sim N\left(z_t \hat{\beta}_{t-1}, H_t + z_t \sum\nolimits_{t|t-1} . z_t'\right), \tag{A9}$$

the process of recursive forecasting can be completed.

In the case of multi-model, for model k, $\sum_{t|t-1}^{(k)}, \hat{\theta}_t^{(k)}$ and $\sum_{t|t}^{(k)}$., can be obtained from Eqs. (A5), (A7), and (A8). By changing the format of Eqs. (A2), (A5) and (A6), the equations for a multi-model can be expressed as:

$$\beta_{t-1} \mid L_{t-1} = k, RRV^{t-1} \sim N\left(\hat{\beta}_{t-1}^{(k)}, \sum\nolimits_{t-1|t-1}^{(k)} .\right), \tag{A10}$$

$$\Theta_t \mid L_t = k, RRV^{t-1} \sim N\left(\hat{\theta}_{t-1}^{(k)}, \sum\nolimits_{t|t-1}^{(k)} .\right), \tag{A11}$$

$$\Theta_t \mid L_t = k, RRV^{t} \sim N\left(\hat{\theta}_{t}^{(k)}, \sum\nolimits_{t|t}^{(k)} .\right). \tag{A12}$$

The transition matrix is specified implicitly by using forgetting factors. The approximations involve two forgetting factors λ and α, $(0 \le \alpha \le 1, 0 \le \lambda \le 1)$. By using λ and α in the model prediction, Markov Chain Monte Carlo (MCMC) is no longer needed for drawing transitions between different models. When a new observation is available, according to the derivation of Kalman filtering, the analogous result for DMA is:

$$p(\Theta_{t-1} RRV^{t-1}) = \sum_{k=1}^{K} p(\beta_{t-1}^{(k)} L_{t-1} = k, RRV^{t-1}) Pr(L_{t-1} = k, RRV^{t-1}), \tag{A13}$$

where $p(\beta_{t-1}^{(k)} \mid L_{t-1} = k, RRV^{t-1})$ is given by Eq. (40). Now let $\pi_{t|s,l} = Pr\,(L_t = l \mid RRV^s)$, and then $Pr(L_{t-1} = k, RRV^{t-1})$ can be written as $\pi_{t-1|t-1,k}$. For the unrestricted transition matrix, P, the elements p_{kl} will be:

$$\pi_{t|t-1,k} = \sum_{l=1}^{K} \pi_{t-1|t-1,l} p_{kl,} \tag{A14}$$

Again, following the approach of Raftery et al. (2012), by involving forgetting factor α $(0 \le \alpha \le 1)$, which is used for the parameters in DMA framework, Eq. (A14) can be written as:

$$\pi_{t|t-1,k} = \frac{\pi_{t-1|t-1,k}^{\alpha}}{\sum_{l=1}^{K} \pi_{t-1|t-1,l}^{\alpha}}. \tag{A15}$$

Thus, the updating equation can be given as:

$$\pi_{t,k} = \frac{\pi_{t|t-1,k} p_k(y_t \mid y^{t-1})}{\sum_{l=1}^{K} \pi_{t|t-1,l} p_l(y_t \mid y^{t-1})}, \tag{A16}$$

where$p_l(RRV_t \mid RRV^{t-1})$ is the predictive density for model l evaluated at RRV_t. In the end, using $\pi_{t|t-1,\kappa}$, recursive forecasting can be done by averaging the forecasting result of each HAR-RRV-type model. Therefore, the model-average 1-step ahead forecasting of RRV_t is given by:

$$\widehat{RRV}_t^{DMA} \equiv \mathbf{E}(RRV_t | RRV^{t-1}) = \sum_{\kappa=1}^{K} \pi_{t|t-1,\kappa} z_t^{(\kappa)} \hat{\beta}_{t-1}^{(\kappa)}. \tag{A17}$$

In our case, we set $\lambda = 0.99$ and $\alpha = 0.94$. In the procedure of DMA, the probability of being model k is calculated and used to average the forecasts generated from different models. For the 1-step ahead, DMA approach will allocate a larger weight to the HAR-RRV-type model that has a better performance. As a special case, DMS approach selects the largest value for $\pi_{t|t-1,k}$ of the HAR-RRV-type model and simply uses it for 1-step ahead forecasting. Note that, when α=0, all models are equally weighted; while for $\alpha = 1$, there is no discounting. BMA is as a special case of DMA by setting $\alpha = \lambda = 1$.

Appendix B

For robustness, we evaluate DMA approach by using MCS test with alternative forgetting factor values and using different frequency data. Besides MCS test, alternative evaluation methods such as error statistics and mean mixed statistics (MME), Diebold-Mariano (DM) test are used and provided in this section.

We evaluate DMA by using alternative forgetting factor values with λ = 0. 99, α = 0. 94.

Based on 10- and 15-min high-frequency data, we also examine DMA's performance. Tables B2 and B3 report the MCS tests for 10-min and 15-min high-frequency data, respectively. Table B2 exhibits the results for 10-min data of oil futures and the S & P 500 index. Under *HMSE* and *HMAE*, the p-values of DMA are all significant across different assets and the p-value is the largest for six out of eight values. That is, DMA is still the best model when using 10-min high-frequency data.

Table B3 shows that most of the models perform good based on the measure T_R. While DMA approach can provide p-values larger than 0.25 across different assets under the two loss functions, they are not always the largest as shown for 5-min and 10-min data. That is because DMA approach is more adaptive to changing data with structural breaks, and hence DMA is stronger in capturing the volatility dynamics. Since 5 minutes

Table B1
MCS tests of out-of-sample forecasts (λ = 0. 99, α = 0. 94).

Model	Crude future oil				the S & P500			
	HMSE		*HMAE*		*HMSE*		*HMAE*	
	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}
BMA	0.0000	0.0000	0.0000	0.0000	0.0860	0.0000	0.0181	0.0000
BMS	0.0000	0.0000	0.0000	0.0000	0.0860	0.0000	0.0181	0.0000
DMA	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
DMS	**0.2751**	0.1976	0.0662	0.1874	0.0860	0.0013	0.0181	0.0000
HAR-RRV	0.1350	0.0002	0.0130	0.0000	0.0860	0.0000	0.0181	0.0000
HAR-RRV-J	0.1350	0.1388	0.0130	0.0331	0.0860	0.0000	0.0181	0.0000
HAR-RRV-CJ	0.1350	0.0000	0.0130	0.0000	0.0860	0.0000	0.0181	0.0000
LHAR-RRV-CJ	**0.2751**	0.1976	0.0662	0.0070	0.0860	0.0000	0.0181	0.0000
HAR-RRV-ONI	0.1350	0.0005	0.0130	0.0000	0.0860	0.0000	0.0181	0.0000
M1	0.1350	0.1976	0.0130	0.1874	0.0860	0.0013	0.0181	0.0000
M2	0.1350	0.0055	0.0130	0.0006	0.0860	0.0000	0.0181	0.0000
M3	0.2135	0.1976	0.0130	0.1874	0.0860	0.0027	0.0181	0.0000
M4	0.0000	0.0000	0.0000	0.0000	0.0860	0.0000	0.0181	0.0000
M5	0.0000	0.0000	0.0000	0.0000	0.0860	0.0000	0.0181	0.0000
M6	0.1350	0.1678	0.0130	0.1874	0.0860	0.0000	0.0181	0.0000
GARCH(1,1)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
EGARCH	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Notes: The MCS *p*-values are computed according to the statistics $T_R$and T_{SQ}. The numbers in this table are the MCS p-values for the different forecasts. The larger the number is, the better the corresponding model performs. The p-values, which are larger than 0.25 indicated in bold and under and underline, denote that the corresponding models are the best based on MCS test. M1-M6 represent the combined models: equal-weighted average, the median combination, the trimmed mean, and the shrinkage forecasts for κ = 0.25, 0.5, and 1.

high-frequency data has been proved to be a rule-of-thumb and it is the best trade-off between market microstructure noise and measurement accuracy (e.g., Andersen and Bollerslev, 1998; Corsi et al., 2010; Sévi, 2014; Liu et al., 2015). In our study, we mainly consider the models' performance based on 5-min high-frequency data of crude oil futures and the S & P 500 index. In conclusion, DMA approach shows its strong forecasting ability and it is also robust in forecasting RRV based on different high-frequency data.

The forecasting models are evaluated by using error statistics, mean mixed statistics (MME) besides MCS test in our revised version.

First, we calculate the magnitude of the forecast errors such as ME, MAE, RMSE, and MAPE. Their specifications are given blow:

$$\mathrm{ME}=\frac{1}{N}\sum_{n=1}^{N}(\hat{\sigma}_T^2-\sigma_T^2) \tag{B1}$$

$$\mathrm{MAE}=\frac{1}{N}\sum_{n=1}^{N}|\hat{\sigma}_T^2-\sigma_T^2| \tag{B2}$$

Table B2
MCS tests of out-of-sample forecasts for 10-min high-frequency data.

Model	Crude oil futures				the S & P500			
	HMSE		*HMAE*		*HMSE*		*HMAE*	
	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}
BMA	0.0000	0.0000	0.0000	0.0000	**1.0000**	**1.0000**	0.0000	0.0000
BMS	0.0000	0.0000	0.0000	0.0000	**1.0000**	**1.0000**	0.0000	0.0000
DMA	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**0.4841**	**0.4841**	**1.0000**	**1.0000**
DMS	**0.7251**	**0.6546**	0.1083	0.0371	**0.4724**	0.0731	0.0002	0.0000
HAR-RRV	**0.5801**	0.0005	0.0000	0.0000	**0.3234**	0.0000	0.0000	0.0000
HAR-RV-J	**0.5801**	0.0159	0.0000	0.0000	**0.3234**	0.0000	0.0000	0.0000
HAR-RV-CJ	**0.5801**	0.1222	0.0000	0.0000	**0.3234**	0.0000	0.0000	0.0000
LHAR-RRV-CJ	**0.7251**	**0.6546**	0.1083	0.0371	**0.3234**	0.0000	0.0000	0.0000
HAR-RRV-ONI	**0.5801**	0.0009	0.0000	0.0000	**0.3234**	0.0000	0.0000	0.0000
M1	**0.7251**	**0.6546**	0.0000	0.0000	**0.3234**	0.0731	0.0002	0.0000
M2	**0.5801**	0.0043	0.0000	0.0000	**0.3234**	0.0000	0.0000	0.0000
M3	**0.5801**	**0.6546**	0.0000	0.0000	**0.4724**	0.0731	0.0002	0.0000
M4	**0.5801**	0.1877	0.0002	0.0003	**0.3234**	0.0000	0.0000	0.0000
M5	**0.5801**	0.2486	0.0000	0.0000	**0.3234**	0.0000	0.0000	0.0000
M6	**0.5801**	0.3928	0.0000	0.0001	**0.3234**	0.0000	0.0000	0.0000
GARCH(1,1)	0.0000	0.0000	0.0000	0.0000	**0.3234**	0.0000	0.0000	0.0000
EGARCH	0.0000	0.0000	0.0000	0.0000	**0.3234**	0.0000	0.0000	0.0000

Notes: The MCS *p*-values are computed according to the statistics $T_R$and T_{SQ}. The numbers in this table are the MCS p-values for the different forecasts. The larger the number is, the better the corresponding model performs. The p-values, which are larger than 0.25 indicated in bold and under and underline, denote that the corresponding models are the best based on MCS test. M1-M6 represent the combined models: equal-weighted average, the median combination, the trimmed mean, and the shrinkage forecasts for κ = 0.25, 0.5, and 1.

Table B3
MCS tests of out-of-sample forecasts for 15-min high-frequency data.

Model	Crude oil futures				S & P500			
	HMSE		*HMAE*		*HMSE*		*HMAE*	
	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}
BMA	0.0000	0.0000	0.0000	0.0000	**0.9267**	0.0000	**0.9333**	0.0000
BMS	0.0000	0.0000	0.0000	0.0000	**0.9267**	0.0000	**0.9333**	0.0000
DMA	**0.9213**	**0.7332**	**1.0000**	**1.0000**	**0.9267**	**0.6002**	**0.9333**	**0.8641**
DMS	**0.7132**	0.1189	0.1739	0.1739	**0.9267**	0.0000	**0.7633**	0.0000
HAR-RRV	**0.7132**	0.0013	0.0018	0.0000	**0.7457**	0.0000	**0.7633**	0.0000
HAR-RRV-J	**0.7132**	0.0259	0.0381	0.0000	**0.7457**	0.0000	**0.7633**	0.0000
HAR-RRV-CJ	**0.9213**	**0.4862**	0.0381	0.0207	**1.0000**	**1.0000**	**1.0000**	0.0000
LHAR-RRV-CJ	**0.7132**	0.1189	0.0381	0.0244	**0.7457**	0.0000	**0.7633**	**1.0000**
HAR-RRV-ONI	**0.7132**	0.0031	0.0018	0.0000	**0.9267**	0.0000	**0.9333**	0.0000
M1	**0.9213**	0.1189	0.0381	0.0000	**0.9267**	0.0000	**0.9333**	0.0000
M2	**0.7132**	0.0158	0.0381	0.0000	**0.7457**	0.0000	**0.7633**	**0.8641**
M3	**0.9213**	0.1189	0.0018	0.0000	**0.7457**	0.0000	**0.7633**	0.0000
M4	**0.9213**	**0.7332**	0.0381	0.0244	**0.7457**	0.0000	**0.7633**	0.0000
M5	**1.0000**	**1.0000**	0.0390	0.0417	**0.7457**	0.0000	**0.7633**	0.0000
M6	**0.9213**	**0.7332**	0.0381	0.0244	**0.7457**	0.0000	**0.7633**	0.0000
GARCH(1,1)	**0.7132**	0.0000	0.0018	0.0000	**0.7457**	0.0000	**0.7633**	0.0000
EGARCH	**0.7132**	0.0000	0.0018	0.0000	**0.7457**	0.0000	**0.7633**	0.0000

Notes: The MCS *p*-values are computed according to the statistics $T_R$and T_{SQ}. The numbers in this table are the MCS p-values for the different forecasts. The larger the number is, the better the corresponding model performs. The p-values, which are larger than 0.25 indicated in bold and under and underline, denote that the corresponding models are the best based on MCS test. M1-M6 represent the combined models: equal-weighted average, the median combination, the trimmed mean, and the shrinkage forecasts for κ = 0.25, 0.5, and 1.

$$\text{RMSE}=\sqrt{\frac{1}{N}\sum_{n=1}^{N}(\hat{\sigma}_T^2-\sigma_T^2)^2} \tag{B3}$$

$$\text{MAPE}=\frac{1}{N}\sum_{n=1}^{N}|(\hat{\sigma}_T^2-\sigma_T^2)/\sigma_T^2| \tag{B4}$$

Table B4 shows the error statistics for oil futures price while Table B5 shows the error statistics for the S & P 500 index. From Table B4, DMA approach provides the smallest errors under various error statistics. From Table B5, we can find that the smallest errors are provided by combined models with constant parameter and DMA approach. Among them, DMA approach shows the smallest errors under ME and MAPE. Hence, DMA approach provides more accurate forecasts for both oil futures and the S & P 500 index based on the magnitude of the forecast errors.

Second, to account for the potential asymmetry in the loss function, we apply an error statistic, which penalizes under-predictions more heavily. It is called the mean mixed error, MME(U), and given as follows:

$$\text{MME(U)} = \frac{1}{N}\left[\sum_{n=1}^{O}|\hat{\sigma}_n^2-\sigma_n^2|+\sum_{n=1}^{U}\sqrt{|\hat{\sigma}_n^2-\sigma_n^2|}\right], \tag{B5}$$

Table B4
Error statistics from forecasting RRV (oil futures price).

Model	ME	MAE	RMSE	MAPE
BMA	0.301567	0.544212	0.828023	0.700737
BMS	0.303004	0.545442	0.828834	0.702462
DMA	**0.022187**	**0.396352**	**0.755199**	**0.393964**
DMS	0.024785	0.398829	0.760614	0.396799
HAR-RRV	0.065481	0.427995	0.803543	0.437009
HAR-RRV-J	0.056329	0.418860	0.795360	0.421087
HAR-RRV-CJ	0.075116	0.432209	0.797566	0.452651
LHAR-RRV-CJ	0.062061	0.417772	0.766604	0.411258
HAR-RRV-ONI	0.067523	0.425491	0.800764	0.435090
M1	0.065302	0.413863	0.783217	0.417234
M2	0.064687	0.421605	0.795667	0.429609
M3	0.065302	0.413863	0.783217	0.417234
M4	3.216902	3.243319	3.982148	3.688121
M5	2.166369	2.206025	2.748895	2.549713
M6	0.065302	0.413863	0.783217	0.417234
GARCH(1,1)	−1.097548	1.097548	1.519214	0.999641
EGARCH	−1.097502	1.097502	1.519155	0.999587

Notes: The numbers in the table are the values for four different error statistics across seventeen models. The smallest number for each error statistic is indicated in bold and underline. ME is the mean error defined by Eq. B1; MAE is the mean absolute error defined by Eq. B2; RMSE is the root mean squared error defined by Eq. B3; MAPE is the mean absolute percentage error defined by Eq. B4 (Brailsford and Faff, 1996).

Table B5
Error statistics from forecasting RRV (the S & P 500 index).

Model	ME	MAE	RMSE	MAPE
BMA	-0.000009	0.000021	0.000053	0.821686
BMS	-0.000009	0.000021	0.000053	0.821686
DMA	**0.000001**	0.000018	0.000089	**0.669722**
DMS	**0.000001**	0.000018	0.000090	0.670615
HAR-RRV	0.000005	0.000017	0.000044	0.918692
HAR-RRV-J	0.000009	0.000019	0.000044	1.124103
HAR-RRV-CJ	0.000004	0.000016	0.000043	0.875261
LHAR-RRV-CJ	0.000000	0.000024	0.000044	1.212386
HAR-RRV-ONI	0.000005	0.000017	0.000043	0.896002
M1	0.000005	**0.000014**	**0.000041**	0.690905
M2	0.000006	0.000016	0.000043	0.890415
M3	0.000005	**0.000014**	**0.000041**	0.690905
M4	0.000006	0.000015	0.000042	0.798200
M5	0.000006	0.000015	0.000042	0.797520
M6	0.000006	0.000015	0.000042	0.796229
GARCH(1,1)	0.000049	0.000052	0.000073	3.238174
EGARCH	0.000054	0.000056	0.000079	3.063650

Notes: The numbers in the table are the values for four different error statistics across seventeen models. The smallest number for each error statistic is indicated in bold and underline. ME is the mean error defined by Eq. B1; MAE is the mean absolute error defined by Eq. B2; RMSE is the root mean squared error defined by Eq. B3; MAPE is the mean absolute percentage error defined by Eq. B4 (Brailsford and Faff, 1996).

Table B6
MME from forecasting RRV and results of under- and over-prediction (crude oil futures).

Model	MME(U)	MME(O)	Under-prediction	Over-prediction	Binomial Prob.
BMA	0.539543	0.677687	141	709	0.000000
BMS	0.540634	0.678751	141	709	0.000000
DMA	**0.423014**	**0.515373**	308	542	0.000000
DMS	0.425622	0.516587	307	543	0.000000
HAR-RRV	0.443420	0.547995	278	572	0.000000
HAR-RRV-J	0.437798	0.535730	287	563	0.000000
HAR-RRV-CJ	0.444173	0.557040	264	586	0.000000
LHAR-RRV-CJ	0.455479	0.517136	342	508	0.000000
HAR-RRV-ONI	0.440671	0.544324	277	573	0.000000
M1	0.431052	0.534961	276	574	0.000000
M2	0.437289	0.541954	278	572	0.000000
M3	0.431052	0.534961	276	574	0.000000
M4	3.236728	1.720459	5	845	0.000000
M5	2.196979	1.416192	7	843	0.000000
M6	0.431052	0.534961	276	574	0.000000
GARCH(1,1)	0.9763923	1.0975478	850	0	0.000000
EGARCH	0.9763693	1.0975017	850	0	0.000000

Notes: Calculated values in this table are provided for two error statistics across seventeen models based on the 850 out-of-sample forecasts of crude oil futures. MME(U) is a mean mixed error which penalizes under-predictions more heavily and is defined by Eq. B5. MME(O) is a mean mixed error which penalizes over-predictions more heavily and is defined by Eq. B6. These statistics are designed to capture potential asymmetry in the loss function. The number of under- and over-predictions are provided for each set of forecasts. The associated binomial probability is based on the test that the number of under- and over-predictions are equal.

where O is the number of over-predictions and U is the number of under-predictions. Similarly, the above statistic can be redefined to weight over-predictions more heavily and is given as follows:

$$\mathrm{MME(O)} = \frac{1}{N}\left[\sum_{n=1}^{O}\sqrt{|\hat{\sigma}_n^2 - \sigma_n^2|} + \sum_{n=1}^{U}|\hat{\sigma}_n^2 - \sigma_n^2|\right]. \tag{B6}$$

The mean mixed error results of crude oil futures are exhibited in Table B6, which shows that DMA provides the smallest value for both MME(O) and MME(U) implying that DMA outperforms others. This is consistent with the conclusion from MCS tests. Table B7 shows MME results of the S & P 500 index. DMA provides the smallest value under MME(O). When using MME(U) as a criterion for the S & P 500 index, the performance of DMA is not good as for oil futures price. Generally, DMA approach shows a smaller tendency of under- or over-prediction.

Both DM test and MCS test are used to evaluate volatility models (Ahoniemi and Lanne, 2013). DM test, comparing models in pair, provides another perspective besides MCS test, which was used in the previous version of our manuscript[4]. In DM test, the key hypothesis is that the two models have equal predictive accuracy and it is directly tested by those forecast error loss differentials.

Table B8 reports the results of DM statistics for competing forecasts of DMA versus alternative one. It is obvious that the difference of MSE

[4] For robustness check, several methods are applied for comparing the models' performance. Besides MCS test, error statistics, mean mixed statistics, and DW test are also calculated both for stock market and oil futures market.

Table B7
MME from forecasting RRV and results of under- and over-prediction (the S & P 500 index).

Model	MME(U)	MME(O)	Under-prediction	Over-prediction	Binomial Prob.
BMA	0.002824	0.001050	705	287	0.000000
BMS	0.002824	0.001050	705	287	0.000000
DMA	0.001554	**0.001755**	444	548	0.332398
DMS	0.001555	0.001758	444	548	0.332398
HAR-RRV	0.000987	0.002636	261	731	0.000000
HAR-RRV-J	0.000792	0.003115	201	791	0.000000
HAR-RRV-CJ	0.001056	0.002504	276	716	0.000000
LHAR-RRV-CJ	0.002357	0.002026	566	426	0.051173
HAR-RRV-ONI	0.001045	0.002560	281	711	0.000000
M1	0.000865	0.002383	265	727	0.000000
M2	0.000921	0.002655	251	741	0.000000
M3	0.000865	0.002383	265	727	0.000000
M4	0.002357	0.002026	566	426	0.051173
M5	0.000757	0.002653	216	776	0.000000
M6	0.000757	0.002652	215	777	0.000000
GARCH(1,1)	0.000180	0.006675	24	968	0.000000
EGARCH	**0.000141**	0.006917	15	977	0.000000

Notes: Calculated values in this table are provided for two error statistics across seventeen models based on the 992 out-of-sample forecasts of the S & P 500 index. MME(U) is a mean mixed error which penalizes under-predictions more heavily and is defined by Eq. B5. MME(O) is a mean mixed error which penalizes over-predictions more heavily and is defined by Eq. B6. These statistics are designed to capture potential asymmetry in the loss function. The number of under- and over-predictions are provided for each set of forecasts. The associated binomial probability is based on the test that the number of under- and over-predictions are equal.

Table B8
DM test for competing forecasts of DMA versus alternative model.

Model comparison	Crude oil futures			The S & P 500 index		
	5-min	10-min	15-min	5-min	10-min	15-min
DMA vs BMA	-.7078	-2.89e-06	-3.02e-06	-7.48e-07	-2.87e-06	-7.36e-06
DMA vs BMS	-.7092	-2.89e-06	-3.02e-06	-7.48e-07	-2.87e-06	-7.36e-06
DMA vs DMS	-.6007	-2.89e-06	-3.02e-06	-7.53e-07	-2.88e-06	-7.36e-06
DMA vs HAR-RRV	-.6679	-2.89e-06	-3.02e-06	-7.47e-07	-2.87e-06	-7.36e-06
DMA vs HAR-RRV-J	-.6548	-2.89e-06	-3.02e-06	-7.47e-07	-2.87e-06	-7.36e-06
DMA vs HAR-RRV-CJ	-.6583	-2.89e-06	-3.02e-06	-7.47e-07	-2.87e-06	-7.36e-06
DMA vs LHAR-RRV-CJ	-.6099	-2.89e-06	-3.02e-06	-7.47e-07	-2.87e-06	-7.36e-06
DMA vs HAR-RRV-ONI	-.6634	-2.89e-06	-3.02e-06	-7.47e-07	-2.87e-06	-7.36e-06
DMA vs M1	-.6356	-2.89e-06	-3.02e-06	-7.47e-07	-2.87e-06	-7.36e-06
DMA vs M2	-.6553	-2.89e-06	-3.02e-06	-7.47e-07	-2.87e-06	-7.36e-06
DMA vs M3	-.6356	-2.89e-06	-3.02e-06	-7.47e-07	-2.87e-06	-7.36e-06
DMA vs M4	-15.88	-2.89e-06	-3.02e-06	-7.47e-07	-2.87e-06	-7.36e-06
DMA vs M5	-7.579	-2.89e-06	-3.02e-06	-7.47e-07	-2.87e-06	-7.36e-06
DMA vs M6	-.6356	-2.89e-06	-3.02e-06	-7.47e-07	-2.87e-06	-7.36e-06
DMA vs GARCH(1,1)	-2.330	-2.92e-06	-3.05e-06	-7.51e-07	-2.87e-06	-7.36e-06
DMA vs EGARCH	-2.330	-2.95e-06	-3.07e-06	-7.52e-07	-2.87e-06	-7.36e-06

Notes: The numbers in the table are the MSE difference in DM test. The models compared in pair are listed in the first column.

between DMA approach and other model is negative for both crude oil futures and the S & P 500 index using different high-frequency data. DMA also shows its advantage under DM test. This result is consistent with that of MCS test.

In our first manuscript, we apply MCS test to compare all the models including DMA, since there are several advantages for MCS test. For example, MCS test does not need to set benchmark model and allow for more than one best models. Overall, the conclusion of DM test is consistent with that of MCS test. DMA approach outperforms other models.

Finally, we tested the data using three different estimation windows of 1000, 1100 and 1200 days for crude oil futures. We also tested estimation windows of 4020 and 4120 days for the S & P 500 index. The results for the window size of 1000 days are given in our manuscript. Table B9 shows the MCS tests using estimation window of 1100 and 1200 days for crude oil futures. From Table B9, we can find that DMA approach performs the best under both *HMSE* and *HMAE*. Additionally, the combined models with shrinkage forecasts performs better than individual models. Table B10 shows the MCS tests using estimation window of 4020 and 4120 days for the S & P 500 index. It also shows that combined models relatively perform better than individual models. Hence, DMA significantly outperforms the other models.

Overall, for different estimation window sizes, DMA shows its strong forecasting ability even across different financial assets. Thus, the result is consistent and robust.

Notes: The results are based on crude oil futures using estimation window of 4020 and 4120 days based on the S & P 500 index. Notes: The MCS *p*-values are computed according to the statistics $T_R$and T_{SQ}. The numbers in this table are the MCS p-values for the different forecasts. The larger the number is, the better the corresponding model performs. The p-values, which are larger than 0.25 indicated in bold and under and underline, denote that the corresponding models are the best based on MCS test. M1-M6 represent the combined models: equal-weighted average, the median combination, the trimmed mean, and the shrinkage forecasts for κ = 0.25, 0.5, and 1.

Table B9
MCS tests of out-of-sample forecasts using different window size (Crude oil futures).

Model	Estimation window of 1100 days				Estimation window of 1200 days			
	HMSE		*HMAE*		*HMSE*		*HMAE*	
	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}
BMA	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
BMS	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
DMA	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
DMS	**0.4891**	**0.4666**	**0.5041**	**0.4215**	**0.2751**	0.1976	0.0662	**0.2751**
HAR-RRV	0.2005	0.0002	0.0238	0.0000	0.1350	0.0002	0.0130	0.1350
HAR-RRV-J	**0.4890**	0.1384	**0.5041**	0.0284	0.1350	0.1388	0.0130	0.1350
HAR-RRV-CJ	0.2005	0.0000	0.0238	0.0000	0.1350	0.0000	0.0130	0.1350
LHAR-RRV-CJ	**0.4890**	0.2141	**0.5041**	0.0995	0.2751	0.1976	0.0662	0.2751
HAR-RRV-ONI	0.2005	0.0006	**0.5041**	0.0001	0.1350	0.0005	0.0130	0.1350
M1	**0.4890**	0.0515	**0.5041**	0.0001	0.1350	0.1976	0.0130	0.1350
M2	0.2005	0.0052	0.0238	0.0000	0.1350	0.0055	0.0130	0.1350
M3	0.2005	0.0284	0.0238	0.0000	0.2135	0.1976	0.0130	0.2135
M4	**0.4890**	**0.4666**	**0.5041**	**0.4215**	0.0000	0.0000	0.0000	0.0000
M5	**0.4891**	**0.4666**	**0.5041**	**0.4215**	0.0000	0.0000	0.0000	0.0000
M6	**0.4891**	**0.4666**	**0.5041**	**0.4215**	0.1350	0.1678	0.0130	0.1350
GARCH(1,1)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
EGARCH	0.0000	0.0000	0.0238	0.0000	0.0000	0.0000	0.0000	0.0000

Notes: The results are based on crude oil futures using estimation window of 1100 and 1200 days. The MCS *p*-values are computed according to the statistics T_R and T_{SQ}. The numbers in this table are the MCS p-values for the different forecasts. The larger the number is, the better the corresponding model performs. The p-values, which are larger than 0.25 indicated in bold and under and underline, denote that the corresponding models are the best based on MCS test. M1-M6 represent the combined models: equal-weighted average, the median combination, the trimmed mean, and the shrinkage forecasts for κ = 0.25, 0.5, and 1.

Table B10
MCS tests of out-of-sample forecasts using different window size (The S & P500 index).

Model	Estimation window of 4020 days				Estimation window of 4120 days			
	HMSE		*HMAE*		*HMSE*		*HMAE*	
	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}
BMA	0.0882	0.0000	0.0533	0.0000	0.0860	0.0000	0.0181	0.0000
BMS	0.0882	0.0000	0.0533	0.0620	0.0860	0.0000	0.0181	0.0000
DMA	**0.5843**	**0.7887**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
DMS	**0.5759**	**0.3166**	**0.4048**	0.2164	0.0860	0.0013	0.0181	0.0000
HAR-RRV	0.0882	0.0000	0.0533	0.0000	0.0860	0.0000	0.0181	0.0000
HAR-RRV-J	0.0882	0.0000	0.0533	0.0000	0.0860	0.0000	0.0181	0.0000
HAR-RRV-CJ	0.0882	0.0000	0.0533	0.0000	0.0860	0.0000	0.0181	0.0000
LHAR-RRV-CJ	0.0882	0.0000	0.0533	0.0000	0.0860	0.0000	0.0181	0.0000
HAR-RRV-ONI	0.0882	0.0000	0.0533	0.0000	0.0860	0.0000	0.0181	0.0000
M1	**0.5843**	**0.7887**	**0.4048**	0.2164	0.0860	0.0013	0.0181	0.0000
M2	0.0882	0.0000	0.0533	0.0000	0.0860	0.0000	0.0181	0.0000
M3	**1.0000**	**1.0000**	**0.4048**	0.2164	0.0860	0.0027	0.0181	0.0000
M4	**0.5759**	0.0000	0.0533	0.0000	0.0860	0.0000	0.0181	0.0000
M5	**0.5759**	0.0000	0.0533	0.0000	0.0860	0.0000	0.0181	0.0000
M6	**0.5759**	0.0001	0.0533	0.0000	0.0860	0.0000	0.0181	0.0000
GARCH(1,1)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
EGARCH	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

See Tables B1,B6 and B7.

References

Agnolucci, P., 2009. Volatility in crude oil futures: a comparison of the predictive ability of GARCH and implied volatility models. Energy Economics 31 (2), 316–321.

Ahoniemi, K., Lanne, M., 2013. Overnight stock returns and realized volatility. International Journal of Forecasting 29 (4), 592–604.

Aloui, C., Jammazi, R., 2009. The effects of crude oil shocks on stock market shifts behaviour: a regime switching approach. Energy Economics 31 (5), 789–799.

Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International economic review 39 (4), 885–905.

Andersen, T.G., Bollerslev, T., Diebold, F.X., 2007. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. The Review of Economics and Statistics 89 (4), 701–720.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71 (2), 579–625.

Aye, G., Gupta, R., Hammoudeh, S., Kim, W.J., 2015. Forecasting the price of gold using dynamic model averaging. International Review of Financial Analysis 41, 257–266.

Bandi, F.M., Russell, J.R., 2008. Microstructure noise, realized variance, and optimal sampling. The Review of Economic Studies 75 (2), 339–369.

Bekaert, G., Hoerova, M., 2014. The VIX, the variance premium and stock market volatility. Journal of Econometrics 183 (2), 181–192.

Bollerslev, T., Patton, A.J., Quaedvlieg, R., 2016. Exploiting the errors: A simple approach for improved volatility forecasting. Journal of Econometrics 192 (1), 1–18.

Brailsford, T.J., Faff, R.W., 1996. An evaluation of volatility forecasting techniques. Journal of Banking & Finance 20 (3), 419–438.

Çelik, S., Ergin, H., 2014. Volatility forecasting using high frequency data: Evidence from stock markets. Economic modelling 36, 176–190.

Charles, A., Darné, O., 2014. Volatility persistence in crude oil markets. Energy Policy 65, 729–742.

Cheong, C.W., 2009. Modeling and forecasting crude oil markets using ARCH-type

models. Energy policy 37 (6), 2346–2355.
Christensen, K., Podolskij, M., 2006. Range-based estimation of quadratic variation. International Conference on High-Frequency Finance, Technical Report, No. 2006, 37.
Christensen, K., Podolskij, M., 2007. Realized range-based estimation of integrated variance. Journal of Econometrics 141 (2), 323–349.
Christensen, K., Podolskij, M., 2012. Asymptotic theory of range-based multipower variation. Journal of Financial Econometrics 10 (3), 417–456.
Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 7 (2), 174–196.
Corsi, F., Pirino, D., Reno, R., 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. Journal of Econometrics 159 (2), 276–288.
Cremers, K.M., 2002. Stock return predictability: A Bayesian model selection perspective. Review of Financial Studies 15 (4), 1223–1249.
Duong, D., Swanson, N.R., 2015. Empirical evidence on the importance of aggregation, asymmetry, and jumps for volatility prediction. Journal of Econometrics 187 (2), 606–621.
Efimova, O., Serletis, A., 2014. Energy markets volatility modelling using GARCH. Energy Economics 43, 264–273.
Graefe, A., Armstrong, J.S., Jones, R.J., Cuzán, A.G., 2014. Combining forecasts: An application to elections. International Journal of Forecasting 30 (1), 43–54.
Granger, C.W., Hyung, N., 2004. Occasional structural breaks and long memory with an application to the S & P 500 absolute stock returns. Journal of empirical finance 11 (3), 399–421.
Hamilton, J.D., 1983. Oil and the macroeconomy since World War II. The Journal of Political Economy 91 (2), 228–248.
Hansen, P.R., 2005. A test for superior predictive ability. Journal of Business & Economic Statistics 23 (4), 365–380.
Hansen, P.R., Lunde, A., 2005. A forecast comparison of volatility models: does anything beat a GARCH (1, 1)? Journal of applied econometrics 20 (7), 873–889.
Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79 (2), 453–497.
Hendry, D.F., Clements, M.P., 2004. Pooling of forecasts. The Econometrics Journal 7 (1), 1–31.
Huang, Z., Liu, H., Wang, T., 2016. Modeling long memory volatility using realized measures of volatility: A realized HAR GARCH model. Economic Modelling 52, 812–821.
Jarque, C.M., Bera, A.K., 1987. A test for normality of observations and regression residuals. International Statistical Review 55 (2), 163–172.
Kilian, L., 2006. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. The American Economic Review 99 (3), 1053–1069.
Kilian, L., Park, C., 2009. The impact of oil price shocks on the us stock market. International Economic Review 50 (4), 1267–1287.
Koop, G., Korobilis, D., 2012. Forecasting Ination Using Dynamic Model Average. International Economic Review 53, 867–886.
Koop, G., Potter, S., 2004. Forecasting in dynamic factor models using Bayesian model averaging. The Econometrics Journal 7 (2), 550–565.
Koopman, S.J., Jungbacker, B., Hol, E., 2005. Forecasting daily variability of the S & P 100 stock index using historical, realised and implied volatility measurements. Journal of Empirical Finance 12 (3), 445–475.
Lean, H.H., Smyth, R., 2015. Testing for weak-form efficiency of crude palm oil spot and future markets: new evidence from a GARCH unit root test with multiple structural breaks. Applied Economics 47 (16), 1710–1721.
Liu, C., Maheu, J.M., 2008. Are there structural breaks in realized volatility? Journal of Financial Econometrics 6 (3), 326–360.
Liu, L.Y., Patton, A.J., Sheppard, K., 2015. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. Journal of Econometrics 187 (1), 293–311.
Ljung, G.M., Box, G.E., 1978. On a measure of lack of fit in time series models. Biometrika 65 (2), 297–303.
Ma, F., Liu, L., Liu, Z., Wei, Y., 2015. Forecasting realized range volatility: a regime-switching approach. Applied Economics Letters 22 (17), 1361–1365.
Martin, M., Dick, V.D., 2007. Measuring volatility with the realized range. Journal of Econometrics 138 (1), 181–207.
McAleer, M., Medeiros, M.C., 2008. Realized volatility: A review. Econometric Reviews 27 (1-3), 10–45.
Mohammadi, H., Su, L., 2010. International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models. Energy Economics 32 (5), 1001–1008.
Naser, H., 2016. Estimating and forecasting the real prices of crude oil: A data rich model using a dynamic model averaging (DMA) approach. Energy Economics 56, 75–87.
Nomikos, N., Andriosopoulos, K., 2012. Modelling energy spot prices: Empirical evidence from NYMEX. Energy Economics 34 (4), 1153–1169.
Raftery, A.E., Kárný, M., Ettler, P., 2010. Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. Technometrics 52 (1), 52–66.
Raftery, A.E., Kárný, M., Ettler, P., 2012. Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. Technometrics 52 (1), 52–66.
Rossi, B., Inoue, A., 2012. Out-of-sample forecast tests robust to the choice of window size. Journal of Business & Economic Statistics 30 (3), 432–453.
Sévi, B., 2014. Forecasting the volatility of crude oil futures using intraday data. European Journal of Operational Research 235 (3), 643–659.
Santos, D.G., Ziegelmann, F.A., 2014. Volatility Forecasting via MIDAS, HAR and their Combination: An Empirical Comparative Study for IBOVESPA. Journal of Forecasting 33 (4), 284–299.
Stock, J.H., Watson, M.W., 2004. Combination forecasts of output growth in a seven-country data set. Journal of Forecasting 23 (6), 405–430.
Taylor, N., 2007. A note on the importance of overnight information in risk management models. Journal of Banking & Finance 31 (1), 161–180.
Todorova, N., Souček, M., 2014. The impact of trading volume, number of trades and overnight returns on forecasting the daily realized range. Economic Modelling 36, 332–340.
Tseng, T.-C., Chung, H., Huang, C.-S., 2009. Modeling jump and continuous components in the volatility of oil futures. Studies in Nonlinear Dynamics & Econometrics 13 (3), 1–30.
Tseng, T.-C., Lai, H.-C., Lin, C.-F., 2012. The impact of overnight returns on realized volatility. Applied Financial Economics 22 (5), 357–364.
Wang, Y., Ma, F., Wei, Y., Wu, C., 2016. Forecasting realized volatility in a changing world: A dynamic model averaging approach. Journal of Banking & Finance 64, 136–149.
Wei, Y., Wang, Y., Huang, D., 2010. Forecasting crude oil market volatility: further evidence using GARCH-class models. Energy Economics 32 (6), 1477–1484.
White, H., 2000. A reality check for data snooping. Econometrica 68 (5), 1097–1126.
Wright, J.H., 2008. Bayesian Model Averaging and exchange rate forecasts. Journal of Econometrics 146 (2), 329–341.