Journal of Banking & Finance 64 (2016) 136–149

Contents lists available at ScienceDirect

Journal of Banking & Finance

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

Forecasting realized volatility in a changing world: A dynamic model averaging approach

Yudong Wang [a,*], Feng Ma [b], Yu Wei [b], Chongfeng Wu [c]

[a] School of Economics and Management, Nanjing University of Science and Technology, China
[b] School of Economics and Management, Southwest Jiao Tong University, China
[c] Antai College of Economics and Management, Shanghai Jiao Tong University, China

ARTICLE INFO

Article history:
Received 25 November 2014
Accepted 23 December 2015
Available online 31 December 2015

JEL classification:
C58
C22
G14
G13

Keywords:
S&P 500 index
Realized volatility
Dynamic model averaging
Time-varying parameters
Portfolio

ABSTRACT

In this study, we forecast the realized volatility of the S&P 500 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. A dynamic model averaging (DMA) approach is used to combine the forecasts of the individual models. Our empirical results suggest that DMA can generate more accurate forecasts than individual model in both statistical and economic senses. Models that use time-varying parameters have greater forecasting accuracy than models that use the constant coefficients. The superiority of time-varying parameter models is also found in volatility density forecasting.

1. Introduction

Accurate forecasts of volatility are crucial for portfolio optimization, derivatives pricing, and risk management. Recent studies have documented that (intra-daily) high-frequency data are useful for forecasting volatility. High-frequency volatility models measure the so-called realized volatility (RV), a concept that was pioneered by Andersen and Bollerslev (1998). This volatility measure is the sum of the squared intra-daily returns. RV takes advantage of the available intra-daily information and is much less noisy than the conventional volatility measure, which is based on squared daily returns. Therefore, it is not surprising that academics have made much effort to improve the methods for forecasting realized volatility (Andersen et al., 2001a,b, 2003, 2005, 2007; Corsi, 2009; Deo et al., 2006; Koopman et al., 2005; Patton and Sheppard, 2013).

Among the realized volatility models, the heterogeneous autoregressive RV (HAR-RV) model proposed by Corsi (2009) is one of the most popular. Although the specification of HAR-RV is simple, it can capture "stylized facts" in financial market volatility such as long memory and multiscaling behavior. Today, HAR-RV has become the standard benchmark for analyzing and forecasting financial volatility dynamics (e.g., Bandi et al., 2013; Chen et al., 2010; Haugom et al., 2014; Sévi, 2014; Lahaye and Shaw, 2014). Building on Corsi's (2009) work, some extensions of HAR-RV have been developed based on different decompositions of realized volatility. For example, Andersen et al. (2007) decompose realized volatility into continuous sample path and jump components to investigate the contribution of jumps to realized volatility; their HAR-RV-CJ model is based on this decomposition. Corsi et al. (2010) introduce a modified measure of jump, the threshold jump (TJ), in their HAR-RV-TCJ model. Previous studies of realized volatility dynamics have identified the "leverage effect," that is the fact that negative equity returns eventually lead to higher future volatility (see, e.g., Bollerslev et al., 2006; Barndorff-Nielsen et al., 2010; Corsi and Reno, 2012; Chen and Ghysels, 2011). To investigate the role of the "leverage effect" in realized

* Corresponding author at: Xiaolinwei Street 200, Xuanwu District, Nanjing, China. Tel.: +86 13681663442.

E-mail addresses: wangyudongnj@126.com (Y. Wang), mafeng575@126.com (F. Ma), weiyusy@126.com (Y. Wei), cfwu@sjtu.edu.cn (C. Wu).

http://dx.doi.org/10.1016/j.jbankfin.2015.12.010

volatility dynamics, Patton and Sheppard (2013) decompose realized volatility into positive and negative semi-variances based on the signs of intra-daily returns. They construct a measure of a signed jump (SJ), which is equivalent to the difference between the positive semi-variance and the negative semi-variance. Furthermore, they extend HAR-RV to account for the "leverage effect" by adding these signed realized measures to the predictive model.

In this study, we forecast the realized volatility of the S&P 500 index using the HAR-RV model and its extensions. We contribute to the literature in five dimensions. First, unlike most existing studies, which use realized volatility models with constant coefficients, 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 (e.g., Banerjee and Urga, 2005; Granger and Hyung, 2004; Lamourex and Lastrapes, 2010; Liu and Maheu, 2008; Rapach and Strauss, 2008) or switches between different regimes (Calvet and Fisher, 2004; Hamilton and Susmel, 1994). 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 three types of volatility models depending on the degree of time-variation 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. The third type of models under consideration is Markov Regime Switching (MRS) ones which can be considered as a midpoint between CC and TVP models. MRS specification allows for volatility dynamics switching between different regimes due to a few structural breaks. As far as we know, there have been no studies in which the realized volatility dynamics are described in a TVP framework. This study will fill this gap.

Second, we also model the time-varying volatility of realized volatility. Corsi et al. (2008) argue that any realized model might be subject to heteroskedastic errors due to the time-varying volatility of the realized volatility estimator. Assessing the relevance of the volatility of realized volatility in modeling and forecasting is thereby very important. However, in the literature, the volatility of realized volatility is always assumed to be constant. The only investigation of its time-varying behavior is in Corsi et al. (2008).

Third, we use a dynamic model averaging 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 (Stock and Watson, 2003, 2004). Choosing the forecasts of an individual model ignores this model uncertainty, understates the forecasting risks, and is likely to result in poor predictions (Hibon and Evgeniou, 2004; Liu and Maheu, 2009). For example, to capture the effects of jump components on realized volatility, Andersen et al. (2007) 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 (Rossi and Sekhposyan, 2011). 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. Models with a history of good predictions receive large weights in the combined future forecasts. DMA is also compared with other popular strategies such as Bayesian model averaging (BMA) and mean forecast combinations (MFC).

Forth, we evaluate forecasting performances not only in statistical sense just like the work in existing studies but also in economic sense. In most studies on forecasting volatility, several loss functions such as mean squared error (MSE) and mean absolute error (MAE) are always employed to quantify how far the volatility forecasts from the true ones. Intuitively, a higher loss function implies that the corresponding forecasts are less accurate. In comparison with statistical accuracy, market investors are more interested in the economic value of volatility forecasts. In other words, they care more about whether they can make money by using these volatility forecasts. For this motivation, we follow the literature by considering a mean–variance utility investor who allocates his or her assets between stock and risk-free Treasury bill, where the optimal weight of stock in the portfolio is ex ante determined by volatility and mean forecasts of stock return (see, e.g., Guidolin and Na, 2006; Neely et al., 2014; Rapach et al., 2010). When constructing portfolio, we use the same historical average forecasts which are widely taken as the benchmark in stock return forecasting literature (Goyal and Welch, 2008; Rapach et al., 2010; Zhu and Zhu, 2013). In this way, the performances of the portfolios uniquely rely on the accuracy of volatility forecasts since the mean forecasts of stock return are the same. To the best of our knowledge, the economic value of *realized* volatility forecasts has been considered in very few of existing studies except the notable work of Fleming et al. (2003).

Up here, we focus on the point forecast of realized volatility. The distribution of the volatility can be of large interest for some financial applications, more than the volatility mean, especially in asset allocation contexts where the volatility dynamics is controlled (e.g., risk control allocation) under different scenarios. For this motivation, we predict the density of realized volatility. This is also our last contribution to the literature. The distribution of RV is early considered in Andersen et al. (2001a,b) but forecasting densities of RV has not been found in existing studies.

Although forecast combinations are extensively used in macroeconomic forecasting, their application to realized volatility forecasting is fairly rare. Liu and Maheu (2009) use a Bayesian model averaging (BMA) approach to combine realized volatility forecasts of HAR-type models with constant coefficient models. Raftery et al. (2010) argue that BMA is restricted to static problems, as it has been found to be computationally infeasible for TVP problems. DMA can overcome this drawback of BMA by using feasible approximations. Moreover, as we discussed above, Liu and Maheu (2009) do not consider time-variation in parameters and the changes in the volatility of realized volatility over time. More recently, Grassi et al. (2014) add different macroeconomic variables including exchange rate and interest rate to the standard HAR-RV and use DMA to combine these HAR-RV models with exogenous variables. Differently, the realized volatility models considered in our paper are main popular extensions of HAR-RV based on different measures of RV or various decompositions of RV. More importantly, we investigate the economic value of volatility forecasts and also forecast distribution of RV. None of these issues are addressed in Grassi et al. (2014) paper.

We use the 5-min intra-daily price data of the S&P 500 index in forecasting analysis. Eight realized volatility models containing HAR-RV and its various extensions are used to generate forecasts. Following four criteria of loss functions are employed to evaluate the models' forecasting performance: mean squared prediction error and mean absolute error for conditional variance (MSE and MAE, respectively) and for standard deviation (MSD and MAD, respectively). We use an advanced model confidence set (MCS)

(Hansen et al., 2011) to examine whether the forecasting losses of different models are significantly different. We find that among three types of individual models, the TVP specifications perform best while the CC ones perform worst out-of-sample, indicating that considering time-variation in the predictive relationships can improve the predictability of realized volatility. DMA forecasts are more accurate than most individual models, indicating that the predictive ability can be improved by allowing for model change over time. Furthermore, we find that DMA performs as well as BMA but better than some other traditional strategies such as mean combination and discount MSPE combination.

To track the source of prediction uncertainty, we use the methodology of Dangl and Halling (2012) by decomposing prediction variance into three terms: expected observational variance, estimation error variance and model uncertainty. We find that the expected observational variance plays the major role prediction uncertainty, consistent with the finding of Dangl and Halling (2012). Both estimation error variance and model uncertainty have important effects on prediction uncertainty.

We use two popular measures that are Sharpe ratio (SR) and certainty equivalent return (CER) to assess the performances of portfolio. We find that DMA forecasts can lead to the portfolio that presents significantly positive return and higher SR and CER than traditional combination strategies, suggesting higher economic value of DMA forecasts. We use expected logarithmic score and continuous rank probability score to evaluate the accuracy of density forecasts. Our findings also suggest that under both of these two criteria DMA can result in more accurate density forecasts than some traditional combinations widely applied in the literature.

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 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 approach to combine the forecasts of the individual models.

2.1. Realized volatility measure

In their pioneering work, Andersen and Bollerslev (1998) 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 $r_{t,j}$:

$$\mathrm{RV}_t = \sum_{j=1}^{M} r_{t,j}^2, \quad t = 1, 2, \ldots T, \tag{1}$$

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

2.2. Modeling realized volatility

In recent years, the heterogeneous autoregressive RV (HAR-RV) of Corsi (2009) has been the most popular RV model. This model accommodates some of the stylized facts found in financial asset return volatility such as long memory and multiscaling behavior. The HAR-RV is simple to implement, as it only contains three explanatory variables: lagged daily realized volatility ($\mathrm{RV}_{d,t}$), lagged weekly realized volatility ($\mathrm{RV}_{w,t}$), and lagged monthly realized volatility ($\mathrm{RV}_{m,t}$). A standard specification of HAR can be written as follows:

$$\mathrm{RV}_{t+1} = \beta_0 + \beta_d \mathrm{RV}_t + \beta_w \mathrm{RV}_{w,t} + \beta_m \mathrm{RV}_{m,t} + \varepsilon_{t+1}, \tag{2}$$

where $\mathrm{RV}_{w,t}$ is the average RV from day $t-4$ to day t and $\mathrm{RV}_{m,t}$ is the average RV from day $t-21$ to t.

Andersen et al. (2007) extend HAR-RV by adding the jump component, and call the extended model the HAR-RV-J model. The specification of HAR-RV-J is,

$$\mathrm{RV}_{t+1} = \beta_0 + \beta_d \mathrm{RV}_t + \beta_w \mathrm{RV}_{w,t} + \beta_m \mathrm{RV}_{m,t} + \beta_{j,d} J_t + \varepsilon_{t+1}, \tag{3}$$

where the jump component $J_t = \max(\mathrm{RV}_t - \mathrm{BPV}_t, 0)$, $\mathrm{BPV}_t = u_1^{-2}\sum_{j=2}^{M}|r_{t,j-1}||r_{t,j}|$ and $u_1 = (2/\pi)^{0.5} = E(|Z|)$. As discussed, RV can be decomposed into continuous sample path and jump components. Based on this decomposition, Andersen et al. (2007) propose the following HAR-RV-CJ model:

$$\begin{aligned}\mathrm{RV}_{t+1} = {} & \beta_0 + \beta_{c,d}\mathrm{CSP}_t + \beta_{c,w}\mathrm{CSP}_{w,t} + \beta_{c,m}\mathrm{CSP}_{m,t} + \beta_{j,d}\mathrm{CJ}_t \\ & + \beta_{j,w}\mathrm{CJ}_{w,t} + \beta_{j,m}\mathrm{CJ}_{m,t} + \varepsilon_{t+1},\end{aligned} \tag{4}$$

where $\mathrm{CJ}_t = I(Z_t > \Phi_\alpha)\cdot[\mathrm{RV}_t - \mathrm{BPV}_t]$, $\mathrm{CSP}_t = I(Z_t \le \Phi\alpha)\cdot \mathrm{RV}_t + I(Z_t > \Phi_\alpha)\cdot \mathrm{BPV}_t$ and $I(\cdot)$ is the indicator function. $\mathrm{CSP}_{w,t}$ ($\mathrm{CJ}_{w,t}$) and $\mathrm{CSP}_{m,t}$ ($\mathrm{CJ}_{m,t}$) are, respectively, the weekly and monthly averages of the continuous sample path and jumps.

Based on the modified jump statistic of Corsi et al. (2010), denoted by TJ_t, we also use the following HAR-RV-TCJ model,

$$\mathrm{RV}_{t+1} = \beta_0 + \beta_{c,d}\mathrm{TC}_t + \beta_{c,w}\mathrm{TC}_{w,t} + \beta_{c,m}\mathrm{TC}_{m,t} + \beta_{j,d}\mathrm{TJ}_t + \varepsilon_t. \tag{5}$$

To capture the role of the "leverage effect" in volatility dynamics, Patton and Sheppard (2013) develop a series of models using signed realized measures. The first model extends the standard HAR-RV by decomposing daily RV into two semi-variances (HAR-RV-RS-I}),

$$\mathrm{RV}_{t+1} = \beta_0 + \beta_d^+ \mathrm{RS}_t^+ + \beta_d^- \mathrm{RS}_t^- + \beta_w \mathrm{RV}_{w,t} + \beta_m \mathrm{RV}_{m,t} + \varepsilon_{t+1}, \tag{6}$$

where $\mathrm{RS}_t^- = \sum_{j=1}^{M} r_{t,j}^2 I(r_{t,j} < 0)$ and $\mathrm{RS}_t^+ = \sum_{j=1}^{M} r_{t,j}^2 I(r_{t,j} > 0)$. The second model (HAR-RV-RS-II) adds a term that interacts the lagged realized variance with an indicator for the negative daily returns, $\mathrm{RV}_t I(r_t < 0)$:

$$\begin{aligned}\mathrm{RV}_{t+1} = {} & \beta_0 + \beta_d^+ \mathrm{RS}_t^+ + \beta_d^- \mathrm{RS}_t^- + \gamma \mathrm{RV}_{d,t} I(r_t < 0) + \beta_w \mathrm{RV}_{w,t} + \beta_m \mathrm{RV}_{m,t} \\ & + \varepsilon_{t+1}.\end{aligned} \tag{7}$$

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

$$\mathrm{RV}_{t+1} = \beta_0 + \beta_{j,d}\mathrm{SJ}_t + \beta_{bv,d} BV_t + \beta_w \mathrm{RV}_{w,t} + \beta_m \mathrm{RV}_{m,t} + \varepsilon_{t+1}, \tag{8}$$

where $\mathrm{SJ}_t = \mathrm{RS}_t^+ - \mathrm{RS}_t^-$.

The last model for the "leverage effect" disentangles the role of the positive and negative jumps (HAR-RV-SJ-II}):

$$\mathrm{RV}_{t+1} = \beta_0 + \beta_{j,d}^+ \mathrm{SJ}_t^+ + \beta_{j,d}^- \mathrm{SJ}_t^- + \beta_{bv,d} BV_t + \beta_w \mathrm{RV}_{w,t} + \beta_m \mathrm{RV}_{m,t} + \varepsilon_{t+1}, \tag{9}$$

where $\mathrm{SJ}_t^+ = \mathrm{SJ}_t \mathrm{I}(\mathrm{SJ}_t > 0)$ and $\mathrm{SJ}_t^- = \mathrm{SJ}_t \mathrm{I}(\mathrm{SJ}_t < 0)$.

Thus, we have eight HAR-type volatility equations to model and forecast realized volatility. A typology of these models is given in Table 1.

2.3. Dynamic model averaging

The realized volatility models expressed by Eqs. (2)–(9) are actually regressions with constant coefficient models. Ma et al. (2015) extend these models by considering regime changes in parameters due to structural breaks.[1] In this study, we extend the

[1] To save space, we do not give detailed description of regime switching HAR models. One can refer to Ma et al. (2015) paper.

Table 1
Typology of the model specifications used in this paper.

Model number	Model name	Reference	Eq. number in this paper
1.	HAR-RV	Corsi (2009)	Eq. (2)
2.	HAR-RV-J	Andersen et al. (2007)	Eq. (3)
3.	HAR-RV-CJ	Andersen et al. (2007)	Eq. (4)
4.	HAR-RV-TCJ	Corsi et al. (2010)	Eq. (5)
5.	HAR-RV-RS-I	Patton and Sheppard (2013)	Eq. (6)
6.	HAR-RV-RS-II	Patton and Sheppard (2013)	Eq. (7)
7.	HAR-RV-SJ-I	Patton and Sheppard (2013)	Eq. (8)
8.	HAR-RV-SJ-II	Patton and Sheppard (2013)	Eq. (9)

models to include time-varying parameter specifications. Suppose that we have K number of models, each of which takes a different set of predictors (in our case $K = 8$). These time-varying parameter RV models can be briefly written as,

$$\mathrm{RV}_t = x_{t-1}^{(k)}\beta_t^{(k)} + \varepsilon_t, \quad \text{and } \beta_t^{(k)} = \beta_{t-1}^{(k)} + \eta_t^{(k)} \tag{10}$$

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.

Theoretically, one could use a stochastic volatility or GARCH specification for $H_t^{(k)}$ (see, e.g., Corsi et al., 2008). However, to do this will greatly add to the computational burden. Alternatively, we follow the suggestion of Koop and Korobilis (2012) in using an Exponentially Weighted Moving Average (EWMA) estimate of $H_t^{(k)}$,

$$\hat{H}_t^{(k)} = \frac{1-\kappa}{1-\kappa^{t-1}}\sum_{j=1}^{t-1}\kappa^j\left(\mathrm{RV}_{t-j} - x_{t-j-1}^{(k)}\hat{\beta}_{t-j}^{(k)}\right)^2 \tag{11}$$

where the decay factor κ is set to be 0.94, according to the suggestion of Riskmetrics (1996).[2]

Let $L_t \in \{1, 2, ..., K\}$ denote which model is used to make a forecast at time t, $\mathrm{B}_t = \left(\beta_t^{((1)')}, \ldots, \beta_t^{(K)'}\right)'$, and $\mathrm{RV}^t = (\mathrm{RV}_1, \ldots, \mathrm{RV}_t)'$. 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 $k = 1,2, \ldots K$. DMA involves in calculating the probability of being model k. Then, these probabilities are used to average the forecasts across the different models. As a special case, DMS selects a single model with the highest value for $\Pr(L_t = k|\mathrm{RV}^{t-1})$. The details on how to estimate DMA with TVP models in Appendix A.

3. Data and preliminary analysis

In this study, we use 5-min price data from the S&P 500 index for the January 2, 1996 to June 24, 2013 period. The data are 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, we obtain high-frequency data for 4280 business days. All of the price data are taken from the Datastream database.

Fig. 1 depicts the evolution of the realized volatility and jump measures. Their descriptive statistics are shown in Table 2. The mean of RV is greater than BPV, CSP, and TC. This makes sense, as the latter three measures do not contain any jump components. All of the series are right skewed except SJ, and they all display positive kurtosis, suggesting they have non-Gaussian distributions. The Jarque–Bera statistic rejects the null hypothesis of a Gaussian distribution for each series, further confirming the fat-tailed distribution.

4. Empirical results

In this section, we first give the in-sample estimation results of the eight volatility models used in study. Then, we evaluate the out-of-sample forecasting performance of DMA in both statistical and economic senses.

4.1. In-sample estimation results

Table 3 shows the OLS estimates of the eight volatility models over the whole sample period along with the t-statistics based on the Newey–West/Bartlett correction, which allows for series correlation up to the order of 5. We can see that the parameter estimates of daily, weekly, and monthly RV in each model are all significant at the 1% level, suggesting strong persistence in the realized volatility dynamics. The coefficients of the continuous sample path (CSP) in the HAR-RV-CJ and HAR-RV-TCJ models are significant, indicating that it plays an important role in the volatility process. The estimation results for the effects of the jump component on realized volatility are mixed, and vary with the model specification. For example, for the HAR-RV-J, HAR-RV-CJ, and HAR-RV-TCJ models, the parameters of the jump components are not significant, except for the parameter estimate of weekly jumps in the HAR-RV-CJ model. In sharp contrast, the coefficients of the signed jumps are all significant in the HAR-RV-SJ-I and HAR-RV-SJ-II models. The coefficient of the negative semi-variance is positive and much greater than that of the positive semi-variance, implying that negative semi-variance contributes more to realized volatility. This finding reveals the existence of a strong "leverage effect." Moreover, we examine the null hypothesis that positive and negative semi-variances have equal predictive power for realized volatility (i.e., $\beta_d^+ = \beta_d^-$) based on these two model specifications. The chi-square statistics of the standard Wald test reject the null hypothesis at the 1% significance level, indicating that the "leverage effect" is significant.[3] The parameter estimates of the signed jumps, $\beta_{j,d}^+$ and $\beta_{j,d}^-$, are both significantly negative, indicating that both positive ("good") and negative ("bad") jumps lead to lower future volatility; however, the estimated value of $\beta_{j,d}^-$ is greater than that of $\beta_{j,d}^+$, confirming the greater effects of "bad" jumps on volatility.

Furthermore, we perform a sub-sample analysis by dividing the whole sample into two sub-samples with equal lengths, and re-estimate the coefficients. The first sub-sample covers the period from January 2, 1996 through February 3, 2005 (Period I) and the second sub-period is from February 4, 1005 to June 24, 2013 (Period II).

Tables 4 and 5 show the estimation results for Period I and Period II, respectively. In these two different periods, we find some consistent patterns in the predictive ability of realized measures.

[2] See Riskmetrics (1996) for the properties of EWMA estimators. Riskmetrics proposes setting κ to be 0.97 for monthly data and 0.94 for daily data. We will also use the alternative value of κ to do robustness check.

[3] We do not report the chi-square statistics in the table, but they are available upon request.

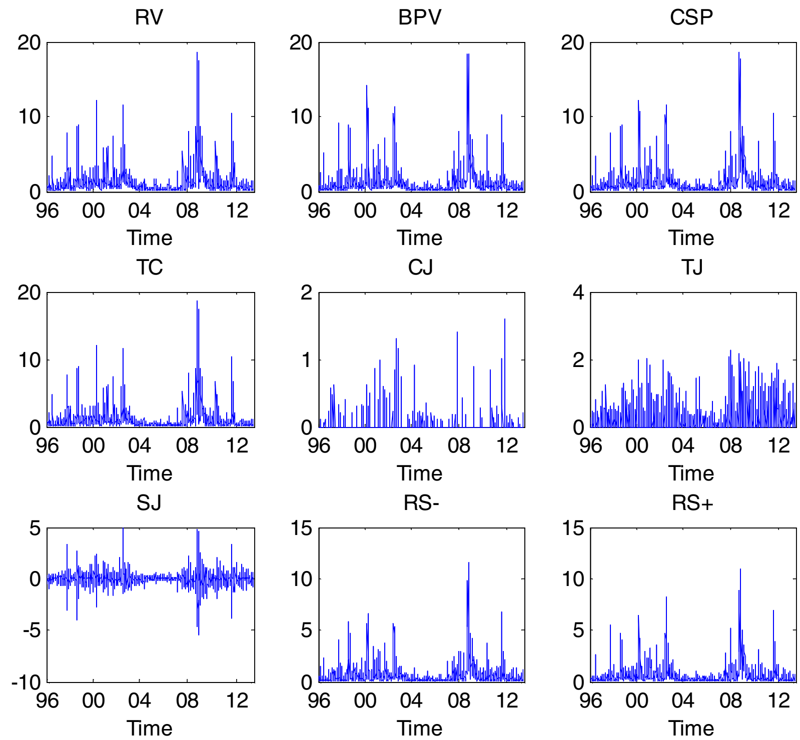

Fig. 1. Evolution of realized measures.

Table 2
Descriptive statistics of realized measures.

	RV	BPV	CSP	TC	CJ	TJ	SJ	RS^-	RS^+
Mean	1.035	0.960	1.026	0.936	0.009	0.099	0.003	0.516	0.519
Median	0.609	0.546	0.595	0.498	0.000	0.000	0.007	0.274	0.292
Maximum	18.62	18.42	18.62	18.62	1.596	2.251	4.908	11.50	10.95
Minimum	0.046	0.037	0.046	0.024	0.000	0.000	−5.454	0.011	0.015
Std. Dev.	1.480	1.448	1.479	1.474	0.072	0.261	0.487	0.798	0.759
Skewness	5.031	5.359	5.054	5.191	12.391	3.786	−0.379	5.268	5.106
Excess kurtosis	35.74	40.58	35.96	37.41	186.84	17.16	24.69	41.62	37.71
Jarque–Bera	2.46e5	3.14e5	2.49e5	2.69e5	6.33e6	6.27e4	1.09e5	3.29e5	2.72e5

Notes: This table reports the descriptive statistics of realized measures. All the Jarque–Bera statistics reject the null hypothesis of Normal distribution at 1% significance level.

First, strong persistent behavior can be seen in the significant parameter estimates of lagged daily, weekly, and monthly volatility. Second, the greater "leverage effect" of negative semivariance on future realized volatility is displayed. Third, negative jumps can result in lower future volatility and have stronger effects than positive jumps.

Despite the consistent patterns discussed above, the values, significance, and even the sign of the parameters in the two subsample periods are not, in most cases, consistent. For example, the estimate of $\beta_{j,d}$ in the HAR-RV-J model is significantly negative for Period I but it is not significant for Period II. The coefficient of the positive jump in the HAR-RV-SJ-IV model ($\beta^+_{j,d}$) is significantly negative for Period I, whereas this estimate is positive for Period II. Overall, our sub-sample analysis shows that the regression coefficients vary with the sample period. In other words, the predictive ability of each predictor of realized volatility changes over time. This makes the use of a time-varying parameter model to forecast volatility more appropriate. Moreover, we find that the parameter estimates rely on the model specification. Different models can reveal different properties of volatility dynamics. Therefore, it is highly possible that the predictive content of realized volatility is affected by both the model specification and sample period. This finding motivates us to use DMA to combine the forecasts of all of the models under consideration.

4.2. Out-of-sample forecasting performance

4.2.1. Forecasting performances of individual models

We find that the in-sample predictive relationships are not constant but change over time. 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, because they are more concerned about the model's ability to improve their future performance than its ability to analyze past patterns. To investigate the effects of time-variation on out-of-sample performance, we consider three types of predictive regressions depending on the degree of time-variation in 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. However, a TVP specification might not be very parsimonious because it is less likely that the structural break in volatility dynamics occurs all the time. For this motivation, the third model specification we

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

	HAR-RV	HAR-RV-J	HAR-RV-CJ	HAR-RV-TCJ	HAR-RV-RS-I	HAR-RV-RS-II	HAR-RV-SJ-I	HAR-RV-SJ-II
β_0	1.10e-5***	6.40e-5***	1.12e-5***	1.67e-5***	9.96e-6***	1.01e-5***	1.12e-5***	1.02e-5***
	(4.208)	(4.016)	(4.199)	(6.574)	(3.863)	(3.925)	(4.351)	(3.933)
β_d	0.445***	0.488***						
	(18.18)	(27.69)						
β_w	0.226***	0.309***			0.300***	0.699***	0.319***	0.320***
	(5.765)	(11.87)			(7.682)	(7.658)	(8.211)	(8.281)
β_m	0.218***	0.153***			0.200***	0.210***	0.224***	0.209***
	(5.980)	(7.250)			(5.614)	(5.820)	(6.236)	(5.836)
$\beta_{c,d}$			0.441***	0.452***				
			(17.96)	(18.51)				
$\beta_{c,w}$			0.233***	0.226***				
			(5.893)	(5.865)				
$\beta_{c,m}$			0.220***	0.230***				
			(5.888)	(6.368)				
$\beta_{j,d}$		−0.124	0.358	0.047			−0.385***	
		(−1.526)	(1.491)	(0.680)			(−10.10)	
$\beta_{j,w}$			−1.112*					
			(−1.845)					
$\beta_{j,m}$			0.936					
			(0.802)					
β_d^+					0.015	0.037		
					(0.305)	(0.735)		
β_d^-					0.795***	0.692***		
					(19.11)	(9.396)		
γ						0.054*		
						(1.699)		
β_{bv}								0.377***
								(16.27)
$\beta_{j,d}^+$								−0.153**
								(−2.100)
$\beta_{j,d}^-$								−0.666***
								(−8.708)

Notes: This table provides the parameter estimation results for the realized volatility models for the whole sample period from January 2, 1996 through June 24, 2013. The numbers in the parentheses are the *t*-statistics. The asterisks *, ** and *** denote rejections of null hypothesis at 10%, 5% and 1% significance levels, respectively.

Table 4
Estimation results of realized volatility models for the first sub-sample.

	HAR-RV	HAR-RV-J	HAR-RV-CJ	HAR-RV-TCJ	HAR-RV-RS-I	HAR-RV-RS-II	HAR-RV-SJ-I	HAR-RV-SJ-II
β_0	5.05e-6**	1.18e-5***	5.43e-6**	1.07e-5***	5.52e-6***	5.47e-6***	6.83e-6***	4.52e-6**
	(2.278)	(4.453)	(2.335)	(4.788)	(2.585)	(2.583)	(3.181)	(1.981)
β_d	0.509***	0.454**						
	(20.71)	(18.32)						
β_w	0.340***	0.225***			0.385***	0.371***	0.421***	0.415***
	(9.586)	(5.720)			(11.20)	(10.83)	(12.35)	(12.35)
β_m	0.106***	0.230***			0.110***	0.118***	0.121***	0.113***
	(3.962)	(6.237)			(4.250)	(4.607)	(4.652)	(4.418)
$\beta_{c,d}$			0.511***	0.514***				
			(20.79)	(20.89)				
$\beta_{c,w}$			0.335***	0.332***				
			(9.452)	(9.457)				
$\beta_{c,m}$			0.109***	0.118***				
			(4.095)	(4.435)				
$\beta_{j,d}$		−0.305**	0.103	0.369***			−0.393***	
		(−2.401)	(0.354)	(5.714)			(−11.94)	
$\beta_{j,w}$			−0.338					
			(−0.523)					
$\beta_{j,m}$			0.597					
			(0.539)					
β_d^+					0.053	0.157***		
					(1.203)	(3.340)		
β_d^-					0.852***	0.626***		
					(23.24)	(12.03)		
γ						0.132***		
						(6.047)		
β_{bv}							0.425***	0.356***
							(17.81)	(14.05)
$\beta_{j,d}^+$								0.029
								(0.441)
$\beta_{j,d}^-$								−0.778***
								(−12.83)

Notes: This table provides the parameter estimation results for the realized volatility models for the first sub-sample period from January 2, 1996 through February 3, 2005. The numbers in the parentheses are the *t*-statistics. The asterisks *, ** and *** denote rejections of null hypothesis at 10%, 5% and 1% significance levels, respectively.

Table 5
Estimation results of realized volatility models for the second sub-sample.

	HAR-RV	HAR-RV-J	HAR-RV-CJ	HAR-RV-TCJ	HAR-RV-RS-I	HAR-RV-RS-II	HAR-RV-SJ-I	HAR-RV-SJ-II
β_0	6.10e-6***	5.05e-6**	6.57e-6***	1.20e-5***	6.50e-6***	6.42e-6***	7.70e-6***	5.71e-6***
	(3.858)	(2.233)	(3.963)	(7.617)	(4.238)	(4.200)	(5.000)	(3.697)
β_d	0.484***	0.510***						
	(27.83)	(20.30)						
β_w	0.309***	0.340***				0.359***	0.395***	0.392***
	(11.89)	(9.512)				(14.16)	(15.71)	(15.70)
β_m	0.149***	0.107***				0.154***	0.159***	0.150***
	(7.114)	(3.953)				(7.556)	(7.747)	(7.366)
$\beta_{c,d}$			0.483***	0.491***				
			(27.77)	(28.27)				
$\beta_{c,w}$			0.309***	0.300***	0.364***			
			(11.87)	(11.68)	(14.30)			
$\beta_{c,m}$			0.152***	0.163***	0.143***			
			(7.222)	(7.828)	(7.035)			
$\beta_{j,d}$		−0.040	0.225	0.239***			−0.400***	
		(−0.369)	(1.198)	(5.101)			(−16.27)	
$\beta_{j,w}$			−0.720					
			(−1.623)					
$\beta_{j,m}$			0.675					
			(0.850)					
β_d^+					0.028	0.088***		
					(0.869)	(2.593)		
β_d^-					0.837***	0.663***		
					(30.82)	(16.11)		
γ						0.098***		
						(5.627)		
β_{bv}							0.403***	0.343***
							(24.15)	(19.19)
$\beta_{j,d}^+$								−0.058
								(−1.258)
$\beta_{j,d}^-$								−0.746***
								(−15.96)

Notes: This table provides the parameter estimation results for the realized volatility models for the second sub-sample period from February 4, 2005 through June 24, 2013. The numbers in the parentheses are the *t*-statistics. The asterisks *, ** and *** denote rejections of null hypothesis at 10%, 5% and 1% significance levels, respectively.

employ is the Markov Regime Switching (MRS) model, a mid-point between CC and TVP models, which allows for switching of volatility dynamics between different regimes due to a few structural breaks.

In this subsection, we will first investigate the forecasting performance of eight individual models and then consider the performance of model averaging. All of the forecasts of each model are generated using recursive regressions. The initial in-sample data for parameter estimation contains the first 2000 observations (i.e., the period from January 2, 1996 through May 4, 2004). To quantitatively evaluate the forecasting accuracy, we follow the literature by using following four popular loss functions:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\sigma_i^2 - \hat{\sigma}_i^2)^2, \tag{12}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\sigma_i^2 - \hat{\sigma}_i^2|. \tag{13}$$

$$\text{MSD} = \frac{1}{n}\sum_{i=1}^{n}(\sigma_i - \hat{\sigma}_i)^2, \quad \text{and} \tag{14}$$

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|\sigma_i - \hat{\sigma}_i|, \tag{15}$$

where σ_i^2 is the actual RV, $\hat{\sigma}_i^2$ is the forecasts of RV and n is the number of forecasts. MSE and MAE are the mean squared error and mean absolute error of realized variance forecasts, respectively. MSD and MAD are these two forecasting losses for standard deviation, respectively.

We assess the statistical significance of differences in forecasting losses using the model confidence set (MCS) developed by Hansen et al. (2011). Following Hansen et al. (2011), Martens et al. (2009), and Rossi and Fantazzini (2014), we use the confidence level of 90%. This allows us to exclude a model with a p-value smaller than 0.1 from the MCS. In other words, the forecasts of this model are significantly less accurate than the models in the MCS. The p-values are obtained based on 10,000 block bootstraps.

We compare the forecasting performances of 8 predictive regressions with CC, MRS and TVP specifications. Table 6 shows the forecasting results of individual models. We can see that TVP models always generate volatility forecasts that have lower loss functions than MRS and 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 three 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. MRS regressions are not incorporated in MCS under the criteria of MSD and MAD. Therefore, we can conclude that relative performances of three types of models depend on the use of loss criteria but TVP models are better choices than CC and MRS ones. These results highlight the importance of allowing for time-variation in parameters of predictive regressions.

Table 6
Forecasting performances of individual models evaluated by loss functions.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
MSE								
CC	0.870 (0.134)	0.870 (0.108)	0.868 (0.182)	0.862 (0.088)	0.835 (0.048)	0.827 (0.106)	0.837 (0.085)	0.841 (0.111)
MRS	0.801 (0.264)	0.816 (0.183)	0.808 (0.316)	0.845 (0.098)	0.813 (0.064)	0.804 (0.139)	0.813 (0.120)	0.801 (0.181)
TVP	**0.723** (1)	**0.726** (1)	**0.733** (1)	**0.722** (1)	**0.705** (1)	**0.707** (1)	**0.717** (1)	**0.716** (1)
MAE								
CC	0.438 (0.056)	0.438 (0.090)	0.437 (0.059)	0.431 (0.017)	0.429 (0.014)	0.429 (0.029)	0.429 (0.055)	0.431 (0.031)
MRS	0.414 (0.418)	0.412 (0.618)	0.410 (0.896)	0.412 (0.108)	0.410 (0.128)	0.410 (0.186)	0.408 (0.339)	0.410 (0.236)
TVP	**0.406** (1)	**0.407** (1)	**0.409** (1)	**0.395** (1)	**0.395** (1)	**0.397** (1)	**0.399** (1)	**0.398** (1)
MSD								
CC	0.076 (0)	0.076 (0.002)	0.076 (0.007)	0.074 (0.001)	0.074 (0)	0.073 (0)	0.073 (0)	0.075 (0)
MRS	0.071 (0.012)	0.071 (0.021)	0.070 (0.067)	0.071 (0.003)	0.070 (0.001)	0.070 (0.002)	0.069 (0.003)	0.069 (0.003)
TVP	**0.063** (1)	**0.064** (1)	**0.065** (1)	**0.061** (1)	**0.061** (1)	**0.061** (1)	**0.061** (1)	**0.061** (1)
MAD								
CC	0.196 (0)	0.196 (0.004)	0.196 (0.012)	0.193 (0)	0.192 (0)	0.192 (0)	0.193 (0)	0.193 (0)
MRS	0.187 (0.081)	0.186 (0.242)	0.185 (0.496)	0.184 (0.005)	0.184 (0.011)	0.184 (0.038)	0.183 (0.069)	0.184 (0.023)
TVP	**0.182** (1)	**0.183** (1)	**0.183** (1)	**0.176** (1)	**0.177** (1)	**0.178** (1)	**0.178** (1)	**0.177** (1)

Notes: This table provides the forecasting results of 8 individual models under consideration evaluated by four loss functions. The names of Models 1–8 are given in Table 1. The numbers in bolds denote that the corresponding model has the lowest loss function under a specific criterion. The numbers in parentheses are the *p*-values of MCS test based on 10,000 block bootstraps. The numbers with underlines denote that the corresponding models are included in MCS for the confidence of 90%.

4.2.2. Forecasting performance of DMA

It has been well documented in the literature that the predictive ability of a single model is quite unstable but changes over time. For this consideration, a dynamic model averaging (DMA) is imposed on individual models. Fig. 2 plots the weights of eight TVP models. We use the forgetting factor $\alpha = 0.99$. We can see that the weight of each model is time-varying. For example, the weight of Model 4 (HAR-RV-TCJ) is always greater than that of the other models during the 2002–2005 period and during the 2009–2013 period. The weight of Model 3 (HAR-RV-CJ) is relatively low but also increases during some short periods of time.

In order to further investigate how well DMA performs out-of-sample, we also consider some alternative strategies. These strategies including model selection (MS), Bayesian model averaging (BMA) and four traditional forecast combinations are imposed over CC, MRS and TVP models. They have some different properties from DMA. For example, DMA can only be applied in a handful of variables (e.g. 20) and computation increases exponentially. In contrast, one can use BMA and these traditional combination methods when more variables (predictors or forecasts) than time series observations are available. Additionally, we consider a TVP HAR model with all predictors where the parameters are obtained via the standard maximum likelihood estimation (MLE). Specifically, we compare the forecasting accuracy of following strategies:

- DMA over time-varying parameter volatility models (DMA-TVP).
- DMA over constant coefficient volatility models (DMA-CC).
- Dynamic model selection (DMS) over TVP models. 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.
- BMA over TVP models (BMA-TVP).
- BMA over CC models (BMA-CC).
- BMS over TVP models (BMS-TVP).
- Mean forecast combination (MFC). This strategy uses the simple equal-weighted average of forecasts from individual models under consideration. MFC is imposed over all CC, MRS and TVP specifications. We denote these methods by MFC-CC, MFC-MRS and MFC-TVP, respectively.
- Trimmed mean combination (TMC). This method uses the equal-weighted average of forecasts from individual models after trimming the one with the worst past performance. TMC is imposed over all CC, MRS and TVP specifications. We denote these methods by TMC-CC, TMC-MRS and TMC-TVP, respectively.
- Discounted mean squared prediction error (DMSPE). DMSPE uses the weighted average of individual forecasts and the weights are given by:

$$w_{i,t} = \phi_{i,t-1}^{-1} \Bigg/ \sum_{j=1}^{N} \phi_{j,t-1}^{-1}, \tag{16}$$

where $\phi_{i,t} = \sum_{s=1}^{t} \delta^{t-s} (\sigma_t^2 - \hat{\sigma}_t^2)^2$. The parameter δ is the discount factor. When $\delta = 1$, there is no discounting. When $\delta < 1$, the model generating more accurate forecasts recently will be assigned greater weights. We follow Zhu and Zhu (2013) by using $\delta = 1$ and 0.9. Also, for consistency, DMSPE method is imposed over all CC, MRS and TVP models. We denote these DMSPE-based strategies by DMSPE (δ)-CC, DMSPE(δ)-MRS and DMSPE(δ)-TVP, respectively.

- A single HAR-RV model with time-varying parameters but the parameters are obtained via standard maximum likelihood estimation (MLE). We denote this strategy by HAR-RV-MLE.

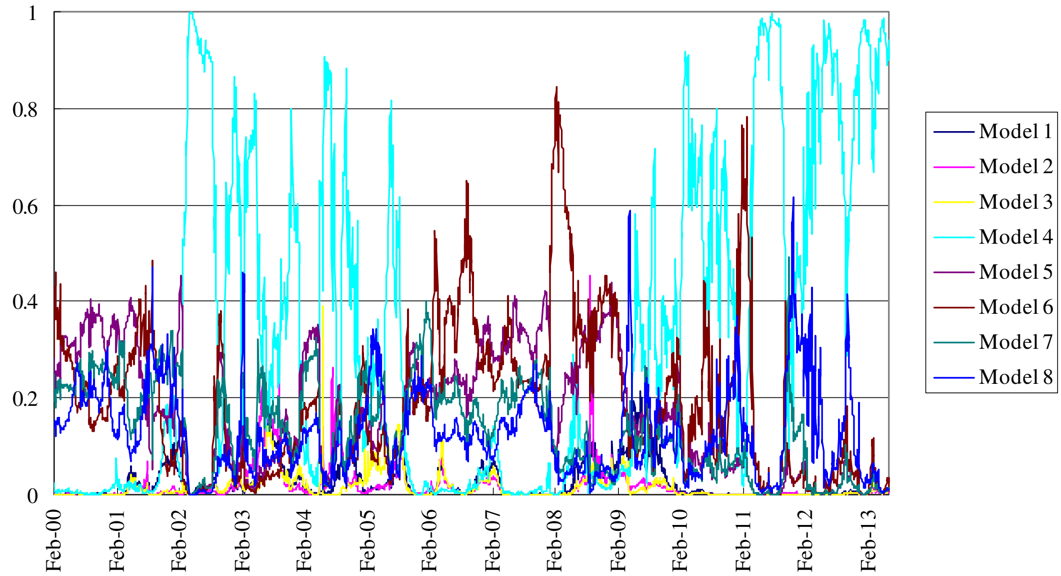

Fig. 2. Weights of volatility models in DMA over time.

In summary, we consider a total of 19 model averaging and forecast combination strategies including the main strategy of DMA-TVP.

Table 7 reports the forecasting performances of abovementioned strategies evaluated by four loss functions, as well as the MCS p-values. Some interesting patterns can be found. First, the third row of Table 7 shows the loss functions of DMA forecasts with TVP regressions. We can find that DMA forecasts have lower forecasting loss than most of individual models shown in Table 6, implying that DMA can improve forecasting accuracy over individual models. Second, DMA forecasts have the lowest loss functions among the 19 strategies under three out of four criteria, indicating the superior forecasting performance. Third, under the criteria of MSE and MAE most of strategies perform as well as DMA-TVP in the statistical sense except a few of combination strategies over CC regressions. Under the other two criteria, most of combination strategies over CC and MRS regressions are excluded from MCS at 90% confidence level, indicating that forecast combinations or model averaging methods over TVP models have significantly better out-of-sample performances. Therefore, the superiority of out-of-sample performances of TVP-based combinations is robust to different loss functions. Forth, the forecasting loss of DMA-TVP is not significantly different from other traditional forecasting combinations over TVP models such as DMSPE-TVP and BMA-TVP. DMA uses a dynamic process to capture time-variation property in the weights of individual models while these traditional combination strategies use the static weights. From our analysis, we can conclude that the time-varying weight cannot significantly improve the volatility forecasting accuracy in the statistical sense. Finally, DMA (BMA) performance is not significantly different from DMS (BMS). Both model averaging and model selection is popular in handling with model uncertainty. Our evidence indicates that the Belmonte and Koop's (2013) question of whether one should choose model averaging or model selection remains unanswered. This also depends on the evaluation criterion.

We have found that DMA-TVP performs well in forecasting realized volatility. Since DMA can track the sources of uncertainty regarding the prediction, we use the methodology of Dangl and Halling (2012) by decomposing the prediction variance of realized volatility into three components. The first component is the expected observational variance. The second component is the expected variance from errors in the estimation of the parameters. This component is always referred to estimation uncertainty. The third component characterizes model uncertainty with respect to variable selection and time-variation of regression coefficients.

We plot the relative weights of these three components of prediction variance over time in Fig. 3. The observational variance (i.e., volatility of volatility) is the dominant source of prediction variance. This finding is not surprising and is consistent with the result in Dangl and Halling (2012). This evidence also highlights the importance of modeling volatility of volatility. The estimation uncertainty in coefficients captures the major fraction of remaining variance but model uncertainty plays more important role in recent several years. Our results from prediction variance decomposition indicate that it is very important for forecasters to reduce estimation error and model uncertainty in forecasting realized volatility.

4.2.3. Robustness test

Two parameters, namely forgetting factors (α and κ), play the key role in a dynamic model averaging process. For robustness check, we investigate the forecasting performances of DMA by using alternative values of α and κ. In our main forecasting analysis, we set $\alpha = 0.99$ and $\kappa = 0.94$. Here, we use $\alpha = 0.95$ following the work of Koop and Korobilis (2012) which implies a more rapid model change. We also use $\kappa = 0.90$ implying a more rapid decay of volatility of volatility with time goes by. The forecasting performances of DMA-TVP based on these alternative parameters are also compared using MCS test. The results shown in Table 8 indicate that the forecasting losses of DMA with different forgetting factors are quite close, regardless of which evaluation criterion is used. MCS results also suggest that these DMA strategies perform equally well in the statistical sense. Also, these DMA strategies have lower forecasting losses than most individual models (see Table 6). Therefore, the predictive ability of DMA is robust to the alternative forgetting factors.

4.3. Portfolio exercise

Comparing with the statistical gains of volatility predictability, market investors care more about economic significance. Specifically, they are interested in how well these volatility forecasts do in asset allocation. To evaluate the economic value of volatility forecasts, we consider a mean–variance utility investor who

Table 7
Forecasting performances of DMA and alternative strategies evaluated by loss functions.

	MSE		MAE		MSD		MAD	
	Loss function	*p*-Value	Loss function	*p*-Value	Loss function	*p*-Value	Loss function	*p*-Value
DMA-TVP	**0.706**	1	0.398	0.820	**0.060**	1	**0.176**	1
DMA-CC	0.839	0.130	0.427	0.224	0.072	0.003	0.191	0.014
BMA-TVP	0.712	0.818	0.396	0.879	0.060	0.773	0.177	0.270
BMA-CC	0.863	0.088	0.430	0.032	0.073	0.001	0.191	0.004
DMS-TVP	0.715	0.796	0.399	0.748	0.061	0.691	0.177	0.210
BMS-TVP	0.713	0.796	0.397	0.832	0.060	0.773	0.176	0.349
MC-TVP	0.716	0.759	0.396	0.832	0.060	0.773	0.177	0.210
TMC-TVP	0.706	0.818	0.396	0.879	0.061	0.773	0.177	0.270
DMSE(1)-TVP	0.706	0.818	0.396	0.879	0.060	0.773	0.177	0.270
DMSE(0.9)-TVP	0.707	0.818	**0.396**	1	0.060	0.773	0.177	0.270
MC-MRS	0.793	0.215	0.405	0.493	0.068	0.029	0.182	0.057
TMC- MRS	0.800	0.150	0.406	0.403	0.069	0.006	0.182	0.048
DMSE(1)- MRS	0.794	0.150	0.405	0.415	0.068	0.006	0.182	0.048
DMSE(0.9)-MRS	0.794	0.150	0.405	0.426	0.068	0.008	0.182	0.084
MC–CC	0.833	0.114	0.428	0.042	0.073	0.001	0.192	0
TMC–CC	0.834	0.098	0.428	0.131	0.073	0.001	0.192	0
DMSE(1)-CC	0.833	0.105	0.428	0.056	0.073	0	0.192	0
DMSE(0.9)-CC	0.834	0.093	0.428	0.081	0.073	0.001	0.192	0
HAR-TVP-ML	0.718	0.775	0.402	0.485	0.063	0.187	0.180	0.062

Notes: This table provides the forecasting performances of DMA and its competing strategies. The numbers in bolds 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 90%.

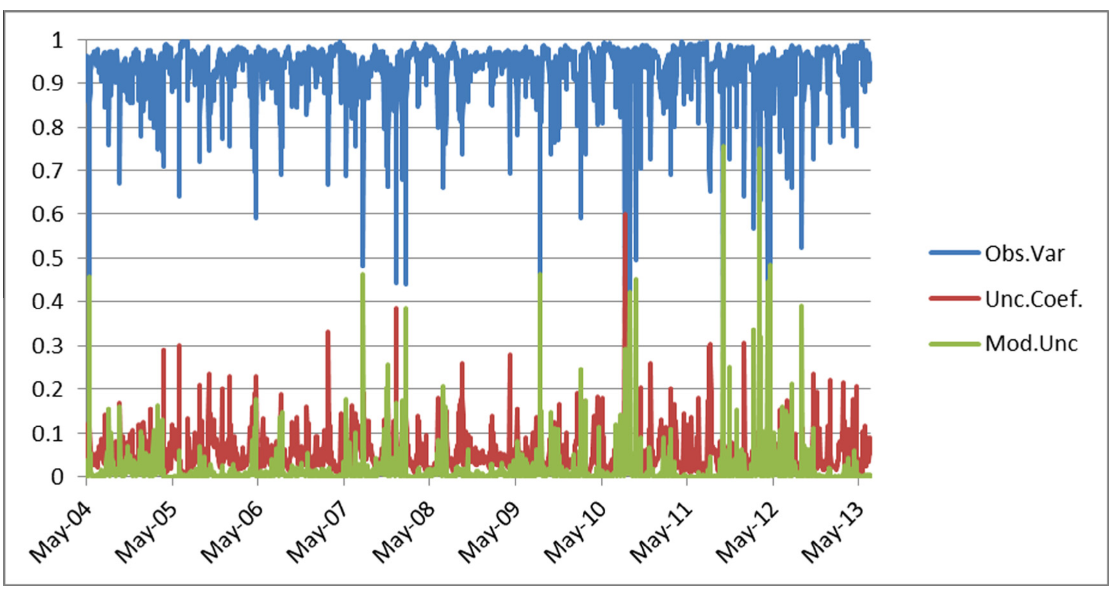

Fig. 3. Weights of three components of prediction uncertainty over time.

allocates his or her assets between stock index and risk-free asset following the literature (see, e.g., Guidolin and Na, 2006; Neely et al., 2014; Rapach et al., 2010).[4] The utility from investing in this portfolio is:

$$U_t(r_t) = E_t(w_t r_t + r_{tf}) - \frac{1}{2}\gamma \, \text{var}_t(w_t r_t + r_{tf}), \tag{17}$$

where w_t is the weight of stock in this portfolio, r_t is the stock return in excess of risk-free rate, r_{tf} is the risk-free rate and γ is the risk aversion coefficient. $E_t(\cdot)$ and $\text{var}_t\,(\cdot)$ denote conditional mean and variance given information at time t.

Maximizing $U_t\,(r_t)$ respect to w_t yield the ex-ante optimal weight of stock index at day $t+1$

$$w_t^* = \frac{1}{\gamma}\left(\frac{\hat{r}_{t+1}}{\hat{\sigma}^2_{t+1}}\right), \tag{18}$$

where $\hat{r}_{t+1}$ and $\hat{\sigma}^2_{t+1}$ are the mean and volatility forecasts of stock excess returns, respectively. We follow the literature by restricting the optimal weight between 0 and 1.5 (i.e., $0 \le w_t^* \le 1.5$) to preclude short sales and preventing more than 50% leverage (Rapach et al., 2010; Neely et al., 2014).

[4] Some papers such as Fleming et al. (2001, 2003) and Guidolin and Timmermann (2005, 2007) also provide some other settings on how to evaluate the economic value of volatility forecasts. They all use volatility forecasts as the key determinant of portfolio optimization by maximizing investor utility and then assess the performances of portfolios formed by volatility forecasts. Therefore, our strategy is intrinsically similar to their settings.

Table 8
Forecasting performances of DMA for alternative priors.

	MSE	MAE	MSD	MAD
DMA (α = 0.99, κ = 0.94)	<u>0.706</u>	<u>0.398</u>	<u>0.060</u>	<u>0.176</u>
	(0.979)	(0.175)	(0.727)	(0.295)
DMA (α = 0.95, κ = 0.94)	**0.705**	**0.396**	**0.060**	**0.176**
	(1)	(1)	(1)	(1)
DMA (α = 0.99, κ = 0.90)	0.706	0.398	0.061	0.177
	(0.945)	(0.175)	(0.727)	(0.295)
DMA (α = 0.95, κ = 0.90)	0.708	0.397	0.060	0.177
	(0.979)	(0.175)	(0.727)	(0.295)

Notes: This table provides the forecasting performances of DMA and its competing strategies for alternative priors. The numbers in bolds denote that the corresponding model has the lowest loss function under a specific criterion. The numbers in parentheses are the *p*-values of MCS test based on 10,000 block bootstraps. The numbers with underlines denote that the corresponding models are included in MCS for the confidence of 90%.

In this way, the portfolio return at day $t+1$ is given by:

$$R_{t+1} = w_t^* r_{t+1} + r_{t+1f}, \tag{19}$$

We employ two popular criteria to evaluate the performance of a portfolio constructed based on return and volatility forecasts. The first is the Sharpe ratio:

$$\mathrm{SR} = \frac{\bar{\mu}_p}{\bar{\sigma}_p}, \tag{20}$$

where $\bar{\mu}_p$ and $\bar{\sigma}_p$ are the mean and standard deviation of portfolio excess returns, respectively.

The second criterion for evaluating portfolio performance is the certainty equivalent return (CER):

$$\mathrm{CER}_p = \hat{\mu}_p - \frac{\gamma}{2}\hat{\sigma}_p^2, \tag{21}$$

where $\hat{\mu}_p$ and $\hat{\sigma}_p^2$ are the mean and variance of portfolio returns over the out-of-sample period, respectively.

It is clear that the optimal weight of stock index in a portfolio depends on the risk aversion degree γ. A higher value of γ implies that stock index is assigned a lower optimal weight in a portfolio. For robustness check, we use three values of 3, 6 and 9 for γ. For the risk-free rate r_{t+1f}, we use the 3-month Treasury bill rate. Actually, the daily risk-free rate is rather close to zero. For the mean forecasts, we use the popular historical average forecasts. HA are generally accepted as the benchmark model in forecasting stock return (see, e.g., Rapach et al., 2010; Neely et al., 2014). Goyal and Welch (2008) find that it is difficult to beat this benchmark in forecasting stock returns out-of-sample. In this way, the optimal weights of stock index is only determined by the volatility forecasts because different strategies share the same mean forecasts of returns when γ is fixed.

Table 9 shows the annualized average mean, Sharpe ratio and CER of portfolios formed by realized volatility forecasts. DMA strategies can result in the portfolio with the annualized return of about 3–6% and the Sharpe ratios of about 0.28–0.35, depending on the value of risk aversion degree γ. For higher values of γ, the mean return from DMA becomes lower but still significantly different from zero. The DMA-TVP portfolio performs as well as BMA-TVP but better than other strategies. The mean returns of portfolios formed by these two strategies are significantly different from zero across all three values of γ, implying their greater performances in the economic sense. More importantly, we find that forecasting combination and model averaging strategies over TVP models generally result in portfolios that have better performances than the strategies over CC and MRS models. This indicates that the use of time-varying parameters in predictive regressions can improve the economic value of realized volatility forecasts.

4.4. Forecasting the density of RV

In this sub-section, we will evaluate the accuracy of volatility mean forecasts under a different environment. We take the mean forecasts of volatility as the key input in constructing the predictive densities and assess how well these forecasts are by comparing the accuracy of density prediction. The distribution of the volatility can be of large interest for some financial applications, more than the volatility mean, especially in asset allocation contexts where the volatility dynamics is controlled (e.g., risk control allocation) under different scenarios.

To measure the accuracy of density forecasts, we use two popular criteria. The first is the expected logarithmic score (lnS) which can be obtained from the average of the sample logarithmic scores:

$$\mathrm{lnS} = -\frac{1}{T}\sum_{t=1}^{T} \ln f_t(\mathrm{RV}_t), \tag{22}$$

where $f_t(\mathrm{RV}_t)$ is the predictive density.

The second criterion is the continuous rank probability score (CRPS) statistic. The CRPS measures the average absolute distance between the empirical cumulative distribution function (CDF) of RV and the empirical CDF that is associated with predictive density:

$$\mathrm{CRPS}_t = \int (F(z) - I\{\mathrm{RV}_t \le z\})^2 dz, \tag{23}$$

where F is the CDF from the predictive density f. $I\{\cdot\}$ is an indicator function which takes the value of 1 when the condition in the brace is satisfied and 0 otherwise. CRPS circumvents some of drawbacks of lnS because the latter one does not reward values from the predictive density that are close but not equal to the realization (see, e.g., Gneiting and Raftery, 2007) and more sensitive to outliers (see, e.g., Groen et al., 2013; Ravazzolo and Vahey, 2014). Higher value of lnS or CRPS implies that the corresponding model results in less accurate density forecasts.

Table 10 shows the performances of DMA-TVP and its competing strategies in the sense of density prediction. Following the literature (see, e.g., Groen et al., 2013), we use the Harvey et al. (1997) small sample correction of Diebold and Mariano (1995) and West (1996) statistics with standard normal critical values to test the null hypothesis that the corresponding strategy and DMA have the equal finite sample forecast accuracy, against the alternative hypothesis that DMA perform better, based on either lnS or CRPS. From Table 10, we can find that DMA-TVP leads to the predictive density with lower lnS or CRPS. It can significantly outperform most of competitors under each of two criteria. Consistently, we also find that forecast combinations based on TVP models perform better than those based on CC and MRS models, further confirming the benefit from time-varying parameters.

5. Conclusions

Intraday high-frequency data is now widely available, and since the seminal work of Andersen and Bollerslev (1998) the modeling and forecasting of realized volatility (RV) has become an important focus of research. In this study, we forecast the realized volatility of the S&P 500 index. Unlike 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 dynamic model averaging (DMA) approach to address this issue of model uncertainty. DMA addresses the problem of the model specification changing over time by combining the forecasts generated from different models. We comprehensively evaluate the forecasting performances in three different dimensions. Our out-of-sample results

Table 9
Performances of portfolios formed by realized volatility forecasts.

	Gamma = 3			Gamma = 6			Gamma = 9		
	R	SR	CER	R	SR	CER	R	SR	CER
DMA-TVP	6.013*	0.321	3.243	4.721*	0.345	2.343	3.381*	0.276	1.613
DMA-CC	5.264	0.264	2.454	3.915	0.237	1.166	2.674	0.144	0.402
BMA-TVP	6.177*	0.331	3.376	4.845*	0.356	2.434	3.443*	0.283	1.640
BMA-CC	5.114	0.253	2.293	3.788	0.223	1.024	2.634	0.138	0.346
DMS-TVP	5.424	0.275	2.599	3.860	0.230	1.096	2.698	0.147	0.401
BMS-TVP	5.109	0.252	2.288	3.653	0.209	0.886	2.556	0.127	0.255
MC-TVP	5.661*	0.296	2.902	4.504*	0.321	2.133	3.307*	0.267	1.566
TMC-TVP	5.658	0.295	2.897	4.552*	0.326	2.170	3.315*	0.266	1.555
DMSE(1)-TVP	5.654	0.295	2.894	4.499*	0.321	2.128	3.303*	0.266	1.562
DMSE(0.9)-TVP	5.654	0.295	2.894	4.504*	0.321	2.129	3.297*	0.265	1.550
MC-MRS	5.908	0.300	2.877	4.701*	0.323	2.024	3.454*	0.272	1.479
TMC-MRS	5.826	0.294	2.796	4.691*	0.322	2.011	3.450*	0.272	1.474
DMSE(1)-MRS	5.900	0.299	2.869	4.696*	0.323	2.020	3.453*	0.272	1.479
DMSE(0.9)-MRS	5.890	0.298	2.859	4.690*	0.322	2.012	3.446*	0.271	1.469
MC-CC	5.665	0.295	2.886	4.328*	0.281	1.601	2.916	0.180	0.687
TMC-CC	5.649	0.294	2.867	4.359*	0.284	1.627	2.958*	0.186	0.722
DMSE(1)-CC	5.667	0.295	2.887	4.328*	0.281	1.600	2.917	0.180	0.689
DMSE(0.9)-CC	5.668	0.295	2.888	4.324*	0.281	1.596	2.916	0.180	0.687
HAR-TVP-ML	5.850*	0.309	3.075	4.437*	0.293	1.713	3.052*	0.199	0.821

Notes: This table shows the performances of portfolios formed by realized volatility forecasts. We give the mean excess return (R), Sharpe ratio and certainty equivalent return (CER) of each portfolio. All the values are annualized. We also test for the null of zero mean of excess return using the standard t-statistic. The asterisk * denote rejection of null hypothesis at 10% significance level.

Table 10
Performances of density forecasts of realized volatility.

	lnS	DM statistic	CRPS	DM statistic
DMA-TVP	0.602	NA	0.435	NA
DMA-CC	0.694***	4.113	0.480***	1.997
BMA-TVP	0.599	−1.201	0.438	0.996
BMA-CC	0.700***	4.377	0.474**	1.737
DMS-TVP	0.607***	2.082	0.437	0.824
BMS-TVP	0.599	−0.915	0.440	1.218
MC-TVP	0.606	0.589	0.438	0.432
TMC-TVP	0.607	0.682	0.437	0.301
DMSE(1)-TVP	0.606	0.588	0.438	0.403
DMSE(0.9)-TVP	0.605	0.533	0.437	0.349
MC-MRS	0.678***	3.465	0.465*	1.437
TMC- MRS	0.678***	3.477	0.466*	1.462
DMSE(1)- MRS	0.678***	3.473	0.465*	1.437
DMSE(0.9)-MRS	0.678***	3.463	0.466*	1.446
MC-CC	0.708***	4.531	0.473**	1.686
TMC-CC	0.707***	4.462	0.474**	1.726
DMSE(1)-CC	0.708***	4.532	0.473**	1.681
DMSE(0.9)-CC	0.707***	4.505	0.473**	1.695
HAR-TVP-ML	0.609***	2.135	0.441	1.245

Notes: This table provides the density forecasting results. We use two criteria, expected logarithmic score (lnS) and the continuous rank probability score (CRPS) to evaluate the density forecasts. We use the Harvey et al. (1997) small sample correction of Diebold and Mariano (1995) and West (1996) statistics with standard normal critical values to test the null hypothesis that the corresponding strategy and DMA have the equal finite sample forecast accuracy. Asterisks *, ** and *** denote rejections of null hypothesis at 10%, 5% and 1% significance levels, respectively.

indicate that time-varying parameter (TVP) models can generate more accurate forecasts than constant coefficient and Markov Regime Switching models in both statistical and economic senses. DMA forecasts also have greater forecasting accuracy than individual models and some traditional forecast combination strategies. These findings indicate the importance of considering parameter change and model specification change in volatility forecasting.

Acknowledgements

We would like to show our sincere gratitude to two anonymous referees whose comments and suggestions greatly improved the quality of this paper. This work is supported by the National Natural Science Foundation of China (Nos. 71501095, 71320107002 and 71371157). Yu Wei is grateful for the financial support from the young scholar fund of science & technology department of Sichuan province (2015JQO010).

Appendix A. Parameter estimation of TVP models

Koop and Korobilis (2012) argue that it is computationally infeasible to estimate TVP models using MCMC-based Bayesian methods. To reduce the calculation burden, we follow Raftery et al. (2010) in using approximations based on two "forgetting factors", α and λ. 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:

$$\beta_{t-1}^{(k)}|\mathrm{RV}^{t-1} \sim N\left(\hat{\beta}_{t-1}^{(k)}, \Sigma_{t-1|t-1}^{(k)}\right)$$

and

$$\beta_{t}^{(k)}|\mathrm{RV}^{t-1} \sim N\left(\hat{\beta}_{t-1}^{(k)}, \Sigma_{t|t-1}^{(k)}\right),$$

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

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

$$\mathrm{RV}_t|\mathrm{RV}^{t-1} \sim N\left(x_{t-1}\hat{\beta}_{t-1}^{(k)}, H_t^{(k)} + x_{t-1}^{(k)}\Sigma_{t|t-1}^{(k)}x_{t-1}^{(k)\prime}\right). \tag{A.1}$$

Here, Raftery et al. (2010) use the forgetting factor method based on the equation,

$$\Sigma_{t|t-1}^{(k)} = \frac{1}{\lambda}\Sigma_{t-1|t-1}^{(k)}, \tag{A.2}$$

or, equivalently, $Q_t^{(k)} = (1-\lambda^{-1})\Sigma_{t-1|t-1}^{(k)}$, where $0 < \lambda \le 1$. 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. In our study, we consider time-varying parameter models with both $\lambda = 0.99$[5] and with a constant coefficient. According to this simplification, one does not need to estimate or simulate $Q_t^{(k)}$; it is only necessary to estimate or simulate $H_t^{(k)}$. Starting with the Kalman filter, we use a diffusion prior $\theta_0^{(k)} \sim N(0, 100I_{m_k})$, where m_k is the number of explanatory variables in model k for $k = 1, \ldots, K$.

After obtaining volatility forecasts based on the individual models, one needs to combine these forecasts using DMA. 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 $\pi_{t|t,k}$[6]. DMA averages the forecasts of the different models using $\pi_{t|t-1,k}$ as weights for $k = 1,\ldots,K$ and $t = 1,\ldots,T$. For instance, the DMA forecasts can be defined as,

$$E(\mathrm{RV}_t|\mathrm{RV}^{t-1}) = \sum_{k=1}^{K} \pi_{t|t-1,k} x_{t-1}^{(k)} \hat{\beta}_{t-1}^{(k)}, \tag{A.3}$$

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

Raftery et al. (2010) use the following equation to describe the relation between $\pi_{t|t-1,k}$ and $\pi_{t-1|t-1,k}$:

$$\pi_{t|t-1,k} = \frac{\pi_{t-1|t-1,k}^{\alpha}}{\sum_{l=1}^{K} \pi_{t-1|t-1,l}^{\alpha}}, \tag{A.4}$$

where $0 < \alpha \leq 1$ is a forgetting factor, a fixed value slightly smaller than one. In this simplification, one does not need an MCMC algorithm to draw transitions between different models; this is important, as such algorithms may cause computational burdens if the number of models is large. Alternatively, we can do a simpler evaluation by updating the equation as follows:

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k} p_k\left(\mathrm{RV}_t|\mathrm{RV}^{t-1}\right)}{\sum_{l=1}^{K} \pi_{t|t-1,l} p_l(\mathrm{RV}_t|\mathrm{RV}^{t-1})}, \tag{A.5}$$

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

$$\begin{aligned} \pi_{t|t-1,k} &\propto [\pi_{t-1|t-2,k} p_k(\mathrm{RV}_{t-1}|\mathrm{RV}^{t-2})]^{\alpha} \\ &= \prod_{i=1}^{t-1} [p_k(\mathrm{RV}_{t-i}|\mathrm{RV}^{t-i-1})]^{\alpha^i}. \end{aligned} \tag{A.6}$$

Therefore, a model that has had a better forecasting performance in the past, as measured by the predictive density $p_k(\mathrm{RV}_{t-i}|\mathrm{RV}^{t-i-1})$, 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. In this study, we follow Koop and Korobilis (2012) in setting $\alpha = 0.99$. We also use an alternative value of $\alpha = 0.95$ to check whether the forecasting accuracy of DMA is robust to a change in this prior value.

Conditional on H_t, the estimation and forecasting strategy described above is only required to evaluate formulae such as those in the Kalman filter. The recursive forecasting procedure is started by choosing $\pi_{0|0,k}$ for $k = 1, \ldots, K$, and we choose $\pi_{0|0,k} = \frac{1}{K}$ as a non-informative prior.

References

Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review 39, 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. Review of Economics and Statistics 89, 701–720.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H., 2001. The distribution of realized stock return volatility. Journal of Financial Economics 61, 43–76.

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

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001. The distribution of exchange rate volatility. Journal of American Statistical Association 96, 42–55.

Andersen, T.G., Bollerslev, T., Meddahi, N., 2005. Correcting the errors: volatility forecast evaluation using high-frequency data and realized volatilities. Econometrica 73, 279–296.

Belmonte, M., Koop, G., 2013. Model switching and model averaging in time-varying parameter regression models. Working paper. University of Strathclyde.

Bandi, F.M., Russell, J.R., Yang, C., 2013. Realized volatility forecasting in the presence of time-varying noise. Journal of Business and Economic Statistics 31, 331–345.

Banerjee, A., Urga, G., 2005. Modelling structural breaks, long memory and stock market volatility: an overview. Journal of Econometrics 129, 1–34.

Barndorff-Nielsen, O.E., Kinnebrock, S., Shephard, N., 2010. Measuring downside risk-realized semivariance. In: Bollerslev, T., Russell, J., Watson, M. (Eds.), Volatility and time series econometrics: essays in honor of Robert F. Engle. Oxford University Press.

Bollerslev, T., Litvinova, J., Tauchen, G., 2006. Leverage and volatility feedback effects in high-frequency data. Journal of Financial Econometrics 4, 353–384.

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.

Chen, X., Ghysels, E., 2011. News-good or bad-and its impact on volatility predictions over multiple horizons. Review of Financial Studies 24, 46–81.

Chen, Y., Hardle, W.K., Pigorsch, U., 2010. Localized realized volatility modeling. Journal of American Statistical Association 105, 1376–1393.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 159, 276–288.

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.

Corsi, F., Mittnik, S., Pigorsch, C., Pigorsch, U., 2008. The volatility of realized volatility. Econometric Reviews 27, 46–78.

Corsi, F., Pirino, D., Reno, R., 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. Journal of Econometrics 159, 276–288.

Dangl, T., Halling, M., 2012. Predictive regressions with time-varying regressions. Journal of Financial Economics 106, 157–181.

Deo, R., Hurvich, C., Lu, Y., 2006. Forecasting realized volatility using a long-memory stochastic volatility model: estimation, prediction and seasonal adjustment. Journal of Econometrics 131, 29–58.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–263.

Fleming, J., Kirby, C., Ostdiek, B., 2001. The economic value of volatility timing. The Journal of Finance 56, 329–352.

Fleming, J., Kirby, C., Ostdiek, B., 2003. The economic value of volatility timing using "realized" volatility. Journal of Financial Economics 67, 473–509.

Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of American Statistical Association 102, 359–378.

Goyal, A., Welch, I., 2008. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21, 1455–1508.

Granger, C.W.J., 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, 399–421.

Grassi, S., Nonejad, N., de Magistris, P.S., 2014. Forecasting with the standardized self-perturbed Kalman Filter. Working paper.

Groen, J.J.J., Paap, R., Ravazzolo, F., 2013. Real-time inflation forecasting in a changing world. Journal of Business and Economic Statistics 31, 29–44.

Guidolin, M., Na, C.F., 2006. The economic and statistical value of forecast combinations under regime switching: an application to predictable US returns. Federal Reserve Bank of St. Louis Working Paper Series, (2006–059).

Guidolin, M., Timmermann, A., 2005. Economic implications of bull and bear regimes in UK stock and bond returns. The Economic Journal 115, 111–143.

Guidolin, M., Timmermann, A., 2007. Asset allocation under multivariate regime switching. Journal of Economic Dynamics and Control 31, 3503–3544.

Hamilton, J.D., Susmel, R., 1994. Autoregressive conditional heteroskedasticity and changes in regime. Journal of Econometrics 64, 307–333.

Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79, 453–497.

Harvey, D., Leybourne, S., Newbold, P., 1997. Testing the equality of prediction mean squared errors. International Journal of Forecasting 13, 281–291.

Haugom, E., Langeland, H., Molnár, P., Westgaard, S., 2014. Forecasting volatility of the US oil market. Journal of Banking & Finance 47, 1–14.

Hibon, M., Evgeniou, T., 2004. To combine or not combine: selecting among forecasts and their combinations. International Journal of Forecasting 21, 15–24.

Koop, G., Korobilis, D., 2012. Forecasting inflation using dynamic model averaging. International Economic Review 53, 867–886.

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, 445–475.

[5] This also follows the suggestion of Raftery et al. (2010).

[6] They begin with $\pi_{0|0,k}$.

Lahaye, J., Shaw, P., 2014. Can we reject linearity in an HAR-RV model for the S&P 500? Insights from a nonparametric HAR-RV. Economic Letters 125, 43–46.
Lamourex, C.G., Lastrapes, W.D., 2010. Persistence in variance, structural change, and the GARCH model. Journal of Business and Economic Statistics 8, 225–234.
Liu, C., Maheu, J.M., 2008. Are there structural break in realized volatility? Journal of Financial Econometrics 6, 326–360.
Liu, C., Maheu, J.M., 2009. Forecasting realized volatility: a Bayesian model-averaging approach. Journal of Applied Econometrics 24, 709–733.
Ma, F., Liu, L., Liu, Z., Wei, Y., 2015. Forecasting realized range volatility: a regime-switching approach. Applied Economics Letters 22, 1361–1365.
Martens, M., Van Dijk, D., De Pooter, M., 2009. Forecasting S&P 500 volatility: long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements. International Journal of Forecasting 25, 282–303.
Neely, C.J., Rapach, D.E., Tu, J., Zhou, G., 2014. Forecasting the equity risk premium: The role of technical indicators. Management Science, forthcoming.
Patton, A.J., Sheppard, K., 2013. Good volatility, bad volatility: Signed jumps and the persistence of volatility. Working paper, Duke University.
Raftery, A., Karny, M., Ettler, P., 2010. Online prediction under model uncertainty via dynamic model averaging: application to a cold rolling mill. Technometrics 52, 52–66.
Rapach, D.E., Strauss, J.K., 2008. Structural breaks and GARCH models of exchange rate volatility. Journal of Applied Econometrics 23, 65–90.
Rapach, D.E., Strauss, J.K., Zhou, G., 2010. Out-of-sample equity premium prediction: combination forecasts and links to the real economy. Review of Financial Studies 23, 821–862.
Ravazzolo, F., Vahey, S.P., 2014. Forecast densities for economic aggregates from disaggregate ensembles. Studies in Nonlinear Dynamics & Econometrics 18, 367–381.
RiskMetrics, 1996. Technical document-forth edition. Available at: <http://www.riskmetrics.com/system/files/private/td4e.pdf>.
Rossi, B., Sekhposyan, T., 2011. Understanding models' forecasting performance. Journal of Econometrics 164, 158–172.
Rossi, E., Fantazzini, D., 2014. Long memory and periodicity in intraday volatility. Journal of Financial Econometrics, nbu006.
Sévi, B., 2014. Forecasting the volatility of crude oil futures using intraday data. European Journal of Operational Research 235, 643–659.
Stock, J.H., Watson, M.W., 2003. Forecasting output and inflation: the role of asset prices. Journal of Economic Perspective 41, 788–829.
Stock, J.H., Watson, M.W., 2004. Combination forecasts of output growth in a seven-country data set. Journal of Forecasting 23, 405–430.
West, K.D., 1996. Asymptotic inference about predictive ability. Econometrica 64, 1067–1084.
Zhu, X., Zhu, J., 2013. Predicting stock returns: A regime-switching combination approach and economic links. Journal of Banking & Finance 37, 4120–4133.