Realized Volatility Forecasting Based on Dynamic Quantile Model Averaging*

Zongwu Cai[a], Chaoqun Ma[b], Xianhua Mi[b,†]

[a]Department of Economics, University of Kansas, Lawrence, KS 66045, USA.

[b]School of Business, Hunan University, Changsha, Hunan 410082, China.

September 25, 2020

Abstract: Heterogeneity, volatility persistence, leverage effect and fat right tails are the most documented stylized features of realized volatility (RV), which introduce substantial difficulties in econometric modeling that requires some rigid distributional assumptions. To accommodate these features without making these assumptions, we study the quantile forecasting of RV by proposing five novel dynamic model averaging strategies designed to combine individual quantile models, termed as dynamic quantile model averaging (DQMA). The empirical results of analyzing high-frequency price data of the S&P 500 index clearly indicate that the stylized facts of RV can be captured by different quantiles, with stronger effects at high-level quantiles. Therefore, DQMA can not only reduce the risk of model uncertainty but also generate more accurate and robust out-of-sample quantile forecasts than those of individual heterogeneous autoregressive quantile models.

Keywords: Dynamic moving averaging; Model uncertainty; Fat tails; Heterogeneity; Quantile regression; Realized volatility; Time-varying parameters.

JEL classification: C12; C13; C14; C23

*The authors acknowledge the financial supports from the National Natural Science Foundation of China (NSFC) key projects with grant numbers 71631004, 72033008 and 71431008, and Emergency Management Project of the NSFC (71850012) as well as China Scholarship Council.

†Correspondence author: Xianhua Mi, E-mail: 939435824@qq.com.

1 Introduction

Accurately acquiring the dynamics of volatility is crucial for many applications in finance. Nevertheless, this task is not trivial due to the established empirical and well-known stylized facts of asymmetry, heavy tails, skewness, discontinuities and long memory, which pose considerable challenges to econometric modeling. In the past two decades, the use of intraday high-frequency data to measure and model volatility has been studied extensively. The so-called realized volatility (RV), which is a model-free nonparametric measurement, first studied by Andersen and Bollerslev (1998), is the most basic and popular measurement. Since then, a burst of academic research has been conducted on modeling RV; see, for example, to name just a few, Barndorff-Nielsen and Shephard (2002), Zhang et al. (2005), Zhang (2006), Christensen et al. (2014), Wang et al. (2016), and Hounyo et al. (2017). Despite these efforts in the literature, at least two interesting factors should be addressed. First, the quantile dynamics of RV are worthy of further study. As a risk measurement, capturing the quantile dynamics of RV is very important in risk management. In fact, to the best of our knowledge, the only work focusing on the quantiles of RV is the paper by Žikeš and Baruník (2016), which applied quantile regression to RV and built a heterogeneous autoregressive quantile model (HARQ). Second, it is important to consider the risks of model uncertainty or/and measuring uncertainty. Although many alternatives for realized measures have been developed, it is still unclear which one is the best under a dynamic changing environment; see, for example, Hendry and Clements (2004) and Aiolfi and Timmermann (2006). In an attempt to address the aforementioned issues, this paper proposes five novel dynamic combination schemes for quantile functions. In the empirical analysis, eleven HARQ models are adopted with more than 35 realized measures of high-frequency data of the S&P 500 Index to comprehensively evaluate their forecasting performances. Our empirical results show clearly that dynamic quantile model averaging (DQMA) can not only reduce the risks of model uncertainty but also generate more accurate and robust out-of-sample quantile forecasts than those of individual HARQ models and other mean models such as dynamic model averaging with time-varying parameter (DMA-TVP) and ARFIMA model.

The main contributions of this paper can be summarized in the following three aspects.

First, in contrast to the existing literature, which focus on modeling the mean of RV, our focus is on forecasting conditional quantiles (distribution), which surely broadens researches of RV. As claimed by Corsi et al. (2008), the distributional assumption for residuals plays a crucial role in density forecasting of RV. Unlike mean regression models, quantile regression has no restrictive assumptions on the conditional distribution, which is efficient and powerful; see, e.g., Fan et al. (1996) and Koenker and Xiao (2004). However, even though quantile regression has been widely used in finance, its applications in RV and other related realized measures are few.

Second, facing risks of model uncertainty, five strategies are proposed for dynamically combining quantile forecasts to model RV under the framework of dynamic model averaging (DMA) and information-theoretic averaging strategies. Though a large quantity of literature on strategies for combining forecasts has been published, which can be dated back to the early work by Bates and Granger (1969) and Granger and Ramanathan (1984), the combinations are usually derived under the framework of the mean squared error (MSE) loss or Bayesian designed mainly for mean regression models, see, e.g., Raftery et al. (2010), Wallis (2011), and the survey papers by Chan and Pauwels (2018), and Steel (2020). Combination strategies designed specifically for quantile estimators are rare, which will be summarized later.

Third, compared to Žikeš and Baruník (2016) by using three HARQ models in their study, eleven HARQ models are adopted, including different realized measures and different components of RV, capturing different RV features, and containing almost all mainstream extensions of the heterogeneous autoregressive RV models in the previous literature. In total, more than 35 realized measures are calculated and used. Therefore, this paper conducts a comprehensive study to provide a framework to accommodate as many realized measures and HARQ models as possible.

Actually, the paper is related to the literature on the stylized features of RV and combination forecasts. On the one hand, it is well documented that the stylized features of RV are extensively studied by the heterogeneous autoregressive RV (HAR-RV) model and its extensions since the paper by Corsi (2009); see, for example, including but not limited to the papers by Corsi et al. (2010), Corsi and Renò (2012), Caporin et al. (2014), Patton and Sheppard (2015), and Audrino and Knaus (2016). However, these previous studies are all

under restrictive distributional (normal) assumptions, which loses sights of its distributional characteristics, see Andersen et al. (2001) for details. To avoid this problem and characterize more features, directly modeling the conditional quantiles of RV via quantile regression seems to be appealing. Quantile regression is more explicable and robust than mean regression when the distribution is skewed or the data contain outliers, see, Koenker and Xiao (2002) and Yu et al. (2003). More details about its good properties and applications can be found in the papers by, to name just a few, Engle and Manganelli (2004), Taylor (2005), and Cai and Xu (2008).

On the other hand, it is well known in empirical applications that using forecasts from a single model inevitably faces risks of model uncertainty so that combination forecasts can outperform individual forecasts; see, for example, Granger and Ramanathan (1984), Hendry and Clements (2004), Stock and Watson (2004), Aiolfi and Timmermann (2006), Liu and Maheu (2009). In fact, Steel (2020) made a full explanation about model uncertainty and stated that "As it is unlikely that reality (certainly in the social sciences) can be adequately captured by any single model, it is often quite risky to rely on a single selected model for inference, forecasts, and conclusion", and Hsiao and Wan (2014) made a detailed description about the reasons and advantages of combination. In fact, this is particularly true for quantile models because a single model is more unlikely to serve all simultaneously when multiple quantile forecasts are of interest, which, however, is put much less attention than mean models. Granger (1989) was the first to consider linearly combining quantile estimators by quantile regression, which was studied extensively by Taylor and Bunn (1998). Also, Giacomini and Komunjer (2005) constructed an encompassing test for comparing conditional quantile forecasts and used the GMM method to obtain the combination weights based on the check loss function. Furthermore, they stated that the combination for value-at-risk (VaR) is beneficial. However, these two papers were not for dynamic combinations. Recently, Bernardi et al. (2017) proposed a dynamic combination technique for VaR predictions using the exponential smoothing moving average process under the framework of GARCH-type models.

To demonstrate on how to capture the above characteristics, this paper conducts comprehensively an empirical study, by employing the quantile dynamics of RV and proposing

five dynamic combination strategies for quantile regressions, intraday high-frequency data of the S&P 500 index price, eleven popular and influential HAR-RV type models, and seven evaluation criteria. Some interesting and new insights are found and summarized as follows.

First, the forecasting performances of DQMA methods are more responsive to the changing dynamic environment and more robust than those of individual models in terms of the risk of model uncertainty. In the empirical study, the results indicate that the performances of DQMA methods are better than or at least equivalent to those of the best individual models. Based on robustness checks, the performance remains almost same if the two best individuals are discarded. Similarly, if the worst model is drooped, the performances of the combination strategies can even be improved. Furthermore, the dynamic evolution of combining weights may also provide a guide for analyzing structural changes by observing the dramatic changes among combining weights.

Second, the stylized facts of RV are different among low-level quantiles and high-level quantiles. For examples, the volatility persistences of high quantiles are stronger than those of low quantiles since lags of RV have larger positive effects on higher quantiles. Similarly, bad volatility and leverage effects are also more impact and stronger on higher conditional quantiles. However, the effects of jumps on the future quantiles of RV are more complex and mixed with no linear trend for different quantiles of RV, though its impacts are almost significant on different quantiles. But the interesting thing is that bad jumps have positive and larger impacts in forecasting high-level quantiles of RV, while good jumps do the opposite. As results, these characteristics also directly lead to stronger predictabilities for high-level quantiles than for low-level quantiles. Also, one can inference from these evidences that the asymmetric effects of volatility persistence, bad volatility, bad jumps and leverage effects between high and low quantiles are powerful factors to produce asymmetric and heavier right tails of RV.

Finally, our empirical studies conclude that quantile models perform much better in modeling the quantile dynamics of RV than mean models such as ARFIMA and DMA-TVP. Indeed, DMA-TVP and ARFIMA can under-forecast the low-level quantiles of RV while over-forecast the high-level quantiles. By comparing the performance of DMA-TVP in Wang et al. (2016), where DMA-TVP outperformed eighteen combining strategies, and the

benchmark status of ARFIMA in modeling volatility, our methods perform very well.

The remainder of this paper is organized as follows. In Section 2, we first introduce the framework of combination quantile regression and then propose our dynamic combination methods designed for quantile functions. Section 3 presents an empirical analysis in detail. In this section, we briefly introduce realized measures and our eleven individual empirical models. Then, seven evaluation criteria are considered. This section also reports the main empirical findings. Robustness checks are performed in Section 4. Finally, Section 5 concludes the paper.

2 Econometric Methodologies

In this section, our dynamic combining approaches are proposed for quantile forecasting models. First, conditional quantile models are briefly introduced together with their linear combinations. Then, the combining strategies of DMA are laid out along with information-theoretic averaging. Finally, our dynamic combining strategies are proposed for quantile forecasting models.

2.1 Conditional Quantile Models and Their Combinations

Let X_t be a vector of predictors and Y_{t+1} be the variable of our interest. Here, X_t may contain some lags of Y_{t+1}. Again, let $F(Y_{t+1}|X_t)$ denote the conditional distribution of Y_{t+1} given X_t. Then, for $0 < \tau < 1$, the τ-th conditional quantile of Y_{t+1} is defined as:

$$q_\tau(X_t) = \inf\{y_{t+1} \in \Re : F(y_{t+1}|X_t) > \tau\} = F^{-1}(\tau|X_t),$$

which may be assumed to have a parametric form as a linear quantile regression $q_\tau(X_t) = \beta_\tau^\top X_t$ as in Koenker and Bassett (1978), where β_τ is the model parameter vector depending on τ. For ease of notation, in what follows, let $q_{t+1,\tau}$ denote $q_\tau(X_t)$. By assuming that several quantile forecasting models, $q_{t+1,\tau,k}$ for $k = 1, 2, \cdots, K$, are available, where $X_{t,k}$ is a vector of predictive variables in the kth model at time t, a linear weighted averaging of

these individual quantile forecasts can be used as a combinational quantile forecast,

$$q_{t+1,\tau} = \sum_{k=1}^{K} w_{t+1|t,\tau,k}\, q_{t+1,\tau,k}, \tag{2.1}$$

where $w_{t+1|t,\tau,k}$ is the dynamic averaging weight given to quantile forecast $q_{t+1,\tau,k}$ at time t. Model (2.1) is based on the simple fact that a linear combination of quantile estimators is still the same quantile estimator if $0 \leq w_{t+1|t,\tau,k} \leq 1$ and $\sum_k w_{t+1|t,\tau,k} = 1$. This is in fact a monotone increasing functional transition of quantiles; see Koenker and Xiao (2006) for details. For the detailed benefits and advantages for a combinational quantile estimator in (2.1), the reader is referred to the related literature such as Yang (2004), Elliott and Timmermann (2004), Hsiao and Wan (2014), and the references therein. Note that the subscript $(t+1)$ in $w_{t+1|t,\tau,k}$ means that weights are given dynamically to $q_{t+1,\tau,k}$ for combining forecast.

Now, the question is how to choose weights in model (2.1). In fact, Granger (1989) proposed to obtaining the combining weights by quantile regression with tick function and constraining the weights to sum to unity, which was applied to empirical investigation by Granger et al. (1989) and studied extensively by Taylor and Bunn (1998). Furthermore, Giacomini and Komunjer (2005) used the GMM method to obtain the combination weights based on the tick loss function (see (2.10) later) after constructing an encompassing test for comparing conditional quantile forecasts. Clearly, these two methods are not dynamic weighting. To do so, Bernardi et al. (2017) was the first to consider dynamically combining quantile models, by proposing a dynamic combination technique for value-at-risk (VaR) predictions using the exponential smoothing moving average process under the framework of GARCH-type models. In the next subsection, dynamic quantile model averaging approaches are considered by following the dynamic moving average strategy in Raftery et al. (2010) and information-theoretic averaging strategies in Burnham and Anderson (2002) and Burnham and Anderson (2004), described next.

2.2 Dynamic Quantile Model Averaging Strategies

The goal of the linear dynamic combination is to find a dynamic weight vector $\mathbf{w}_{t+1|t}$ to form a combined forecast $y^c_{t+1} = \mathbf{w}^\top_{t+1|t}\mathbf{y}^f_{t+1}$ at time t, where $\mathbf{y}^f_{t+1} = \{y^f_{t+1,1}, y^f_{t+1,2}, \cdots, y^f_{t+1,K}\}$, denoting K individual forecasts with predictors $X_{t,k}$. For this purpose, DMA, introduced by Raftery et al. (2010), can be used, defined as follows:

$$y^{\text{DMA}}_{t+1} = \sum_{k=1}^{K} w_{t+1|t,k}\, y^f_{t+1,k}, \tag{2.2}$$

where $y^f_{t+1,k}$ is the forecast of the k-th individual model with predictors $X_{t,k}$ at time t based on a random coefficient linear regression model; see Raftery et al. (2010) for details. Clearly, the choice of $w_{t+1|t,k}$ in (2.2) is the key to the DMA approach.

To obtain the dynamic weight $w_{t+1|t,k}$ in DMA, one needs commonly two components in an updating scheme: $w_{t+1|t,k}$ and $w_{t+1|t+1,k}$. The recursive relationship between $w_{t+1|t,k}$ and $w_{t+1|t+1,k}$ is as follows:

$$w_{t+1|t,k} = \frac{w^{\alpha_0}_{t|t,k}}{\sum_{l=1}^{K} w^{\alpha_0}_{t|t,l}}, \tag{2.3}$$

where $0 < \alpha_0 \leq 1$ is a forgetting factor that is a fixed value and slightly less than one; see Section 3.4.2 on how to choose α_0 in our empirical application. The updating equation is:

$$w_{t+1|t+1,k} = \frac{w_{t+1|t,k} f_k(Y_{t+1}|X_{t,k})}{\sum_{l=1}^{K} w_{t+1|t,l} f_l(Y_{t+1}|X_{t,l})}, \tag{2.4}$$

where $f_k(Y_{t+1}|X_{t,k})$ is the predictive density of the kth model evaluated at Y_{t+1}. Now, the question turns to a challenging issue on how to estimate $f_k(Y_{t+1}|X_{t,k})$ in (2.4), which relies commonly on the assumptions of individual (normal) densities. Usually, this assumption is excessively restrictive and unrealized in practice. Actually, it is difficult to find an appropriate distribution for financial series. However, $f_k(Y_{t+1}|X_{t,k})$ can easily be estimated from a sequence of quantile estimators of τ's without any distributional assumptions under the framework of quantile regression models; see Koenker and Xiao (2004). Once $f_k(Y_{t+1}|X_{t,k})$ is estimated via the conditional quantiles models, it is easy to follow the procedure of DMA.

According to Koenker and Xiao (2004), one can obtain the conditional density of Y_{t+1} given X_t for some appropriately chosen sequence of τ's. Specifically, with $0 < \tau_1 < \cdots <$

$\tau_{n_0+1} < 1$ for some large n_0, one can estimate a sequence of $n_0 + 1$ quantile points $\widehat{q}_{t+1,\tau_i,k}$ at grid point τ_i from the k-th quantile model . Suppose the probability at each quantile interval $\left[\widehat{q}_{t+1,\tau_{i+1},k}, \widehat{q}_{t+1,\tau_i,k}\right]$ is $\tau_{i+1} - \tau_i$, then, the predictive density from the k-th conditional quantile model can be estimated by

$$\widehat{f}_k(Y_{t+1}|X_{t,k}) = \begin{cases} \frac{\tau_{i+1}-\tau_i}{\widehat{q}_{t+1,\tau_{i+1},k,} - \widehat{q}_{t+1,\tau_i,k,}} & \text{if } Y_{t+1} \subseteq \left(\widehat{q}_{t+1,\tau_{i+1},k}, \widehat{q}_{t+1,\tau_i,k}\right) \text{ for } 1 \leq i \leq n_0, \\ 0, & \text{others.} \end{cases} \tag{2.5}$$

Therefore, we can directly get the weights of model (2.1) under the procedure of DMA by replacing the predictive density $f_k(Y_{t+1}|X_{t,k})$ in (2.4) with (2.5). For the initial weights $w_{0|0,k}$, if there is no information prior, equal weights $w_{0|0,k} = 1/K$ should be used. This type of dynamic quantile model averaging strategy is denoted as DQMA-I.

However, the weights in DQMA-I do not depend on τ. To choose the weights depending on τ, one can assign weights subjecting to each conditional quantile via Laplace density. Yu and Moyeed (2001) demonstrated that the minimization of the tick loss function (see (2.10) later) of quantile regression is exactly equivalent to the maximization of a likelihood function formed by combining independently distributed asymmetric Laplace densities defined as follows

$$f_{\tau,k}(u_{t+1,k}) = \tau(1-\tau)\exp\{-\rho_\tau(u_{t+1,k})\}, \tag{2.6}$$

where $u_{t+1,k} = Y_{t+1} - q_{t+1,\tau,k}$ is the predictive error and $\rho_\tau(u) = u(\tau - I(u < 0))$ is the so-called tick loss or check function. Therefore, by replacing the predictive density in (2.4) with this Laplace density (2.6), it is easy to obtain the second-step weights (the weights in (2.4)) for each quantile forecast of the individual models as follows

$$w_{t+1|t+1,\tau,k} = \frac{w_{t+1|t,k} f_{\tau,k}(u_{t+1,k})}{\sum_{l=1}^{K} w_{t+1|t,l} f_{\tau,l}(u_{t+1,l})}, \tag{2.7}$$

which is abbreviated as DQMA-II.

Information-theoretic averaging, which assigns weights based on information, such as the Akaike information criterion (AIC), Akaike information criterion with a correction (AICc) and Bayesian information criterion (BIC), is a popular choice; see Burnham and Anderson (2002) and Burnham and Anderson (2004). The weight for the k-th model based on

information-theoretic averaging is given as follows:

$$w_k^{\text{Info}} = \frac{\exp\left(-0.5\Psi_k^{\text{Info}}\right)}{\sum_{i=1}^{K}\exp\left(-0.5\Psi_i^{\text{Info}}\right)}, \tag{2.8}$$

where $\Psi_k^{\text{Info}} = \Delta_k - \min_{i=1,\cdots,K}\{\Delta_i\}$ and Δ_k is some model fitting information such as AIC, AICc or BIC. Another information-theoretic averaging methods is the inverse of the mean squared error as in Timmermann (2006), where the weights are given by:

$$w_k^{\text{MSE}} = \frac{(\text{MSE}_k)^{-1}}{\sum_{i=1}^{K}(\text{MSE}_i)^{-1}}, \tag{2.9}$$

where MSE_k is the mean squared error for the kth model.

To obtain the weights $w_{t|t,\tau,k}$ for quantile regressions, following the ideas of information-theoretic averaging methods, two approaches are considered. The first is based on the goodness of fit of quantile models. The other is analogous to the MSE method of forecasting the mean of Y_{t+1} based on the predicted density in (2.5). To quantitatively evaluate the forecasting accuracy and obtain the goodness of fit for the quantile models, following the idea from Koenker and Machado (1999), the following check loss function is employed

$$\widehat{L}_\tau = -\sum_{t:y_{t+1}<q_{t+1,\tau}}^{T}(1-\tau)(Y_{t+1}-q_{t+1,\tau}) + \sum_{t:Y_{t+1}\geq q_{t+1,\tau}}^{T}\tau(Y_{t+1}-q_{t+1,\tau}) \tag{2.10}$$

and $\widehat{L}_{0,\tau}$ denotes the check loss function for the case that there is no predictor at all in the model. Then, the quantile goodness of fit is defined as follows

$$\text{Goodness}_\tau = 1 - \widehat{L}_\tau/\widehat{L}_{0,\tau}. \tag{2.11}$$

Hence, one can obtain the weights for the conditional quantile models as:

$$w_{t|t,\tau,k}^{\text{Goodness}} = \frac{\text{Goodness}_{t,\tau,k}}{\sum_{i=1}^{K}\text{Goodness}_{t,\tau,i}}. \tag{2.12}$$

This method is termed as DQMA-Gods.

For the other information-theoretic averaging method related to the MSE information,

two sub-steps are used. First, the fitted mean forecasts of Y_{t+1} are obtained from individual quantile models by

$$\overline{Y}^{q}_{t+1,k} = \sum_{i=1}^{n_0} \left[(\tau_{i+1} - \tau_i) \, \frac{\widehat{q}_{t+1,\tau_{i+1},k} + \widehat{q}_{t+1,\tau_i,k}}{2} \right], \tag{2.13}$$

where $\overline{Y}^{q}_{t+1,k}$ denotes the mean forecast from the k-th conditional quantile model. Then, using these mean forecasts, one can obtain the in-sample MSE loss, $\text{MSE}^{q}_{k} = \frac{1}{T_0} \sum_{t=1}^{T_0} \left(Y_t - \overline{Y}^{q}_{t,k} \right)^2$. Inserting MSE^{q}_{k} into (2.9), another combining method for quantile regression is obtained, which is abbreviated as DQMA-MSE.

However, an often quoted folk theorem in the forecast combination literature is that using equal average as opposed to optimal weights often works better in practice, which is also the well-known phenomenon called *the forecast combination puzzle*. The reason is that it can be difficult to precisely estimate the optimal forecast combination weights. Equal-weighted forecast combinations may be biased but they also reduce the forecast error variance by not relying on estimated combination weights that depend on some statistics of forecast errors. Therefore, equal weights $w^{Eq}_{k} = 1/K$ should be considered as an alternative, denoted by DQMA-Eq. In a sum, together with the equal weighting strategy, five combination strategies are proposed for conditional quantile models based on the ideas of DMA and information-theoretic averaging strategies.

Finally, it should mention the related method recently proposed by Bernardi et al. (2017) for a dynamic combining scheme for VaR under the GARCH-type framework using exponential smoothing moving average process. That is, $w_{t+1,j} = k_j w_{t,j} + (1 - k_j) \pi \left(r_t, \text{VaR}^{\tau}_{t|t-1,j}, \widehat{\sigma}_{t,j} \right)$, where $\pi \left(r_t, \text{VaR}^{\tau}_{t|t-1,j}, \widehat{\sigma}_{t,j} \right)$ is the normalized exponential of the $\tau-$quantile loss kernel, r_t is the return, $\widehat{\sigma}_{t,j}$ is the predicted conditional variance at time t of model j, and k_j is the autoregressive parameter estimated by minimizing the average VaR loss function over the forecast horizon. However, by a comparison, the method in Bernardi et al. (2017) is different from our proposed schemes from the following aspects. First, their method is a form of addition that updates weights via exponential smoothing. By contrast, our method is a form of multiplication; that is, being analogue to the posterior probability for the j-th individual model. Second, they have to estimate k_j under in-sample conditions, which may

induce in-sample over-fitting and leads to a poor forecasting performance.

3 Empirical Analysis

In this section, we investigate RV forecasting based on quantile models and DQMA. First, RV measurement is briefly described and then, our selected eleven HARQ models are introduced. Also, the in-full-sample estimation results of these eleven quantile models are empirically analyzed. Furthermore, the out-of-sample forecasting performances of individual models, DQMA, and two other popular competing models (DMA-TVP and ARFIMA) are evaluated and compared.

3.1 Realized Volatility Measurement

In this subsection, a basic definition of RV and its theory are provided. In fact, a total of 35 realized measures are used in our empirical study. However, to save spaces, the related definitions of other realized measures are skipped. The reader is referred to the related references cited in Subsection 3.3.1 later. Now, consider a continuous-time stochastic process for log-price, p_t, which consists of a continuous component and a jump component,

$$p_t = \int_0^t \mu_s d_s + \int_0^t \sigma_s dw_s + J_t, \tag{3.1}$$

where μ_t is a locally bounded predictable drift process, σ_t is a strictly positive cadlag process, and J_t is a pure jump process. Then, the RV on day t is defined by

$$RV_t = \sum_{j=1}^{M} r_{t,j}^2, \quad 1 \leq t \leq T, \tag{3.2}$$

proposed by Andersen and Bollerslev (1998) as a measure of integrated variance, where $r_{t,j}$ denotes the j-th intraday return as the difference of logarithm price $p_{t,j}$ on day t and M is the number of returns of each day. Indeed, Andersen et al. (2003) showed that RV_t defined in (3.2) converges in probability to the quadratic variation as the time intervals between

observations become infinitely small

$$RV_t \ \rightarrow \ \int_0^t \sigma_s^2 ds + \sum_{0<s\leq t} (\Delta p_s)^2 \tag{3.3}$$

as $M \rightarrow \infty$, where $\Delta p = p_s - p_{s-}$ captures a jump if present. To get ride of the jump part in (3.2), the bi-power variation, introduced by Barndorff-Nielsen and Shephard (2004), can be used

$$BV_t = \frac{\pi}{2} \sum_{i=2}^{M} |r_{t,i}||r_{t,i-1}| \ \rightarrow \ \int_0^t \sigma_s^2 ds \tag{3.4}$$

as $M \rightarrow \infty$, and it was showed by Barndorff-Nielsen and Shephard (2004) that it is a consistent estimator for the continuous component as in (3.3). Clearly, it follows from (3.3) and (3.4) that

$$RV_t - BV_t \ \rightarrow \ \sum_{0<s\leq t} (\Delta p_s)^2,$$

which can be used to capture the jump if present.

3.2 Data and Preliminary Analysis

Following Liu et al. (2015), our focus is on modeling and forecasting RV using 5-minute data. In this study, 10-years price data of the S&P 500 index from October 1, 2008 to September 27, 2018 are adopted. All data come from the trading time of each business day between 9:30:00 and 16:00:00. After some cleaning procedures, we obtain the high-frequency data for 2473 trading days. All the data are taken from the Datastream and Thomson Reuters Database.

First, some preliminary analyses to the data are conducted and Table 1 presents descriptive statistics of all the daily realized measures. The mean of RV is greater than all other continuous component measures (BV, CSP, TC). Moreover, the means of RS^+ and RS^- are greater than CSV^+ and CSV^-, respectively. Note that CSP, TC, RS^+, RS^-, CSV^+, CSV^- and other notations in Table 1 are defined later. This is consistent with our purpose, as the latter measures do not contain any jump components. Additionally, all these measures are right skewed except for negative measures and jumps because they are truncated at zero. All measures have kurtosis greater than three, suggesting that they follow non-Gaussian dis-

tributions. Also, the Jarque-Bera tests are conducted with their statistics not presented in the table, to save spaces. But, they are all large enough to indicate that the null assumption of a normal distribution should be rejected.

Figure 1 shows the evolutions of daily RV_t and other realized measures. Although there are some differences among the realized measurements for the same component, they all have similar behavior, further confirming that the measurements are consistent. The figure also shows that periods of large volatilities are associated with large jumps.

Table 1: Descriptive statistics of all realized measures

	mean	median	min	max	skew	kurtosis	sd	variance
RV	0.6990	0.2757	0.0170	28.5341	8.3515	101.9940	1.5837	2.5081
BV	0.6599	0.2535	0.0174	31.7643	9.2664	126.5340	1.5887	2.5241
CSP	0.6957	0.2739	0.0169	28.5341	8.3074	100.7460	1.5897	2.5271
J	0.0564	0.0165	0.0000	3.5022	9.8754	152.5670	0.1554	0.0242
CJ	0.0068	0.0000	0.0000	1.6186	17.6369	365.3550	0.0633	0.0040
TC	0.6792	0.2661	0.0145	28.5341	8.4259	105.5340	1.5559	2.4209
TJ	0.0234	0.0000	0.0000	12.2217	36.8504	1592.5320	0.2742	0.0752
RS^+	0.3503	0.1374	0.0088	18.0704	9.6267	141.6070	0.8554	0.7316
RS^-	0.3523	0.1348	0.0069	14.9944	7.8997	94.1160	0.7874	0.6199
SJ	-0.0021	-0.0013	-5.2198	7.6067	3.4870	110.2760	0.4088	0.1671
SJ^+	0.0709	0.0000	0.0000	7.6067	13.9906	282.3990	0.3037	0.0922
SJ^-	-0.0730	-0.0013	-5.2198	0.0000	-9.7748	134.1240	0.2540	0.0645
r^-	-0.2716	0.0000	-6.1033	0.0000	-3.9588	23.4650	0.5642	0.3183
$VolJ$	0.0140	0.0000	0.0000	15.1621	40.8980	1847.1890	0.3279	0.1075
CSV^+	0.3257	0.1301	0.0088	15.7947	9.6222	140.4240	0.7894	0.6232
CSV^-	0.3208	0.1225	0.0069	14.9944	8.2052	108.7760	0.7179	0.5154
JSV^+	0.0246	0.0000	0.0000	6.4559	21.2682	578.4560	0.1966	0.0386
JSV^-	0.0316	0.0000	0.0000	8.1230	23.7276	740.6680	0.2253	0.0507
$cRet^-$	-0.2484	0.0000	-6.1033	0.0000	-4.2491	27.5310	0.5370	0.2883
$jRet$	-0.0228	0.0000	-2.8758	3.1528	-0.3618	36.8470	0.2399	0.0576
$jRet^-$	-0.0559	0.0000	-2.8758	0.0000	-5.9035	54.3320	0.1791	0.0321
$cRet$	0.0458	0.0741	-6.1033	6.3238	-0.2073	9.8870	0.8407	0.7067

Note: The return takes the percentage form.

3.3 RV Quantile Forecasting Models

3.3.1 Quantile Models

In this section, eleven HARQ type models are considered by following the framework of Žikeš and Baruník (2016). The general form of HARQ for the conditional τ-quantile of RV_{t+1} given Ω_t is defined as a linear quantile regression model as in Koenker and Bassett (1978), as follows:

$$q_\tau(RV_{t+1}|\Omega_t) = \beta_0(\tau) + \beta_d(\tau)RV_t + \beta_w(\tau)RV_{t-5,t} + \beta_m(\tau)RV_{t-22,t} + \beta_z^\top(\tau)z_t, \tag{3.5}$$

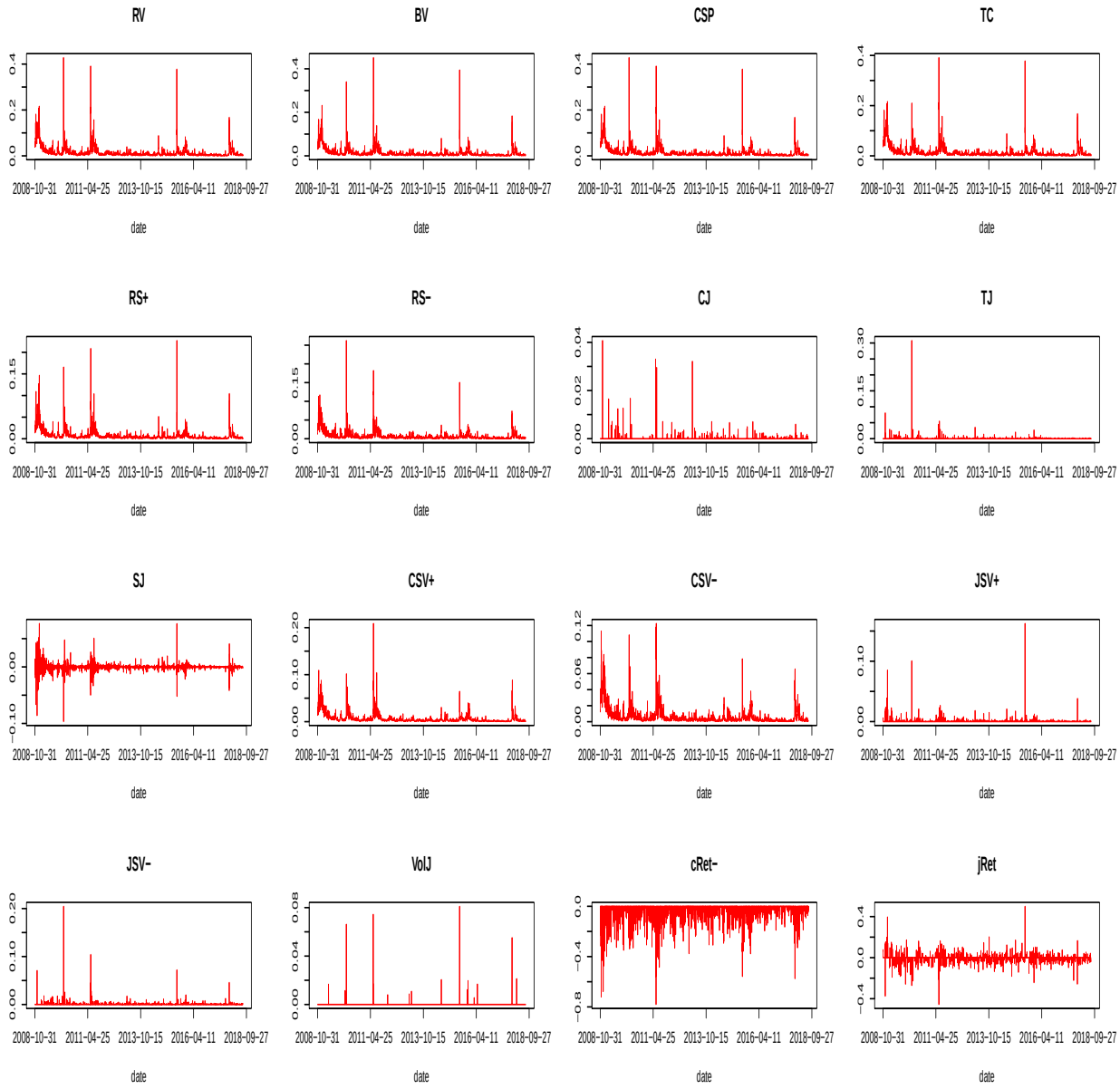

Figure 1: Evolutions of realized measures. The starting date is October 31, 2008, due to the lags of RV for calculating the monthly measure $\text{RV}_{m,t}$.

where Ω_t denotes the information set, which may contain different variables in what it follows; see Models 1 – 11 below, defined in (3.6) – (3.16), $RV_{t-k,t} = \sum_{j=0}^{k-1} RV_{t-j}/k$ for $k = 5$ or 22, z_t is a vector of (weakly) exogenous variables, and $\beta_0(\tau)$, $\beta_d(\tau)$, $\beta_w(\tau)$, $\beta_m(\tau)$ and $\beta_z(\tau)$ are coefficients (vectors) to be estimated. For simplicity, Ω_t is dropped from $q_\tau(RV_{t+1}|\Omega_t)$ and abbreviate $RV_{t-5,t}$ and $RV_{t-22,t}$ as $RV_{w,t}$ and $RV_{m,t}$, which correspond to

working days of a week and a month. Note, all of the realized measurements below with subscripts w and m are defined same with $RV_{w,t}$ and $RV_{m,t}$.

Actually, Žikeš and Baruník (2016) demonstrated the relationship between HARQ and (3.1) in theory and found evidences that the conditional quantile of RV_{t+1} can be approximated by (3.5). As a result, the conditional quantiles of RV_t can be modeled by HARQ without making any distributional assumptions. Also, Žikeš and Baruník (2016) documented that HARQ for RV performs well in capturing the dynamics of the respective conditional distributions. Since Corsi (2009) provided the simple form of HAR-RV model, many extensions have been developed, including but not limited to the papers by Corsi et al. (2010), Corsi and Renò (2012), Caporin et al. (2014), Patton and Sheppard (2015), and Audrino and Knaus (2016).

Model 1: Without z_t in (3.5), our HARQ-RV models is defined as:

$$q_\tau(RV_{t+1}) = \beta_0(\tau) + \beta_d(\tau)RV_t + \beta_w(\tau)RV_{w,t} + \beta_m(\tau)RV_{m,t}, \tag{3.6}$$

which is similar to the model HAR-RV in Corsi (2009).

Model 2: Similar to Andersen et al. (2007), HARQ-RV is extended by adding a jump component to obtain the HARQ-RV-J model, given by

$$q_\tau(RV_{t+1}) = \beta_0(\tau) + \beta_d(\tau)RV_t + \beta_w(\tau)RV_{w,t} + \beta_m(\tau)RV_{m,t} + \beta_{j,d}(\tau)J_t, \tag{3.7}$$

where $J_t = \max[RV_t - BV_t, 0]$ measures the jump.

Model 3: Considering finite sample performances and the empirical phenomenon, the jumps measured in (3.7) exhibit an unreasonably large number of nonzero small positive values, and it is desirable to treat these small jumps as measurement errors. Thus, similar to Andersen et al. (2007), a truncation technique is employed and HARQ-RV-CJ model is given by

$$\begin{aligned} q_\tau(RV_{\mathrm{t+1}}) =& \beta_0(\tau) + \beta_{c,d}(\tau)CSP_{\mathrm{t}} + \beta_{c,w}(\tau)CSP_{w,t} + \beta_{c,m}(\tau)CSP_{m,t} \\ &+ \beta_{j,d}(\tau)CJ_{\mathrm{t}} + \beta_{j,w}(\tau)CJ_{\mathrm{w,t}} + \beta_{j,m}(\tau)CJ_{\mathrm{m,t}}, \end{aligned} \tag{3.8}$$

where CSP_t and CJ_t are the shrinkage estimation of continuous sample path component variation and significant jumps, respectively; see Andersen et al. (2007) for the detailed definitions.

Model 4: To eliminate the small sample bias of RV_t and BV_t, Corsi et al. (2010) proposed an alternative estimator of integrated powers of volatility in the presence of jumps, the threshold bi-power variation, and a test for jump detection in time series. As a result, the HARQ-RV-TCJ model to forecast volatility is given below

$$q_\tau(RV_{t+1}) = \beta_0(\tau) + \beta_{c,d}(\tau)TC_t + \beta_{c,w}(\tau)TC_{w,t} + \beta_{c,m}(\tau)TC_{m,t} + \beta_{j,d}(\tau)TJ_t, \tag{3.9}$$

where TC_t and TJ_t are the corresponding continuous part and jump parts estimated from the threshold bi-power variation, respectively; see Corsi et al. (2010) for details.

Model 5: In empirical applications, the leverage effect is a well-known phenomenon, i.e., that negative returns lead to higher future volatility than positive returns. After studying the impact of signed returns on future volatility for mean regression models, Patton and Sheppard (2015) introduced several forecasting models by using semi-variance estimators introduced by Bollerslev et al. (2010). Here, their four main models are used for our quantile setting and they are described in Models 5 – 8 defined in (3.10) – (3.13), respectively. The first model, denoted as HARQ-RV-RS-I, is obtained by decomposing RV into two semi-variances, defined as follows:

$$q_\tau(RV_{t+1)=\beta_0(\tau)} + \beta_d^+(\tau)RS_t^+ + \beta_d^-(\tau)RS_t^- + \beta_w(\tau)RV_{w,t} + \beta_m(\tau)RV_{m,t}, \tag{3.10}$$

where $RS_t^+ = \sum_{i=1}^{M} r_{i,t}^2 I(r_{i,t} > 0)$ and $RS_t^- = \sum_{i=1}^{M} r_{i,t}^2 I(r_{i,t} < 0)$ are the realized semi-variance estimator, also called good volatility and bad volatility.

Model 6: The second model, termed as HARQ-RV-RS-II, is obtained by adding a term that interacts with the lagged realized variance with an indicator for negative lagged daily

returns, $RV_t I(r_t < 0)$, so that it is defined as

$$q_\tau(RV_{t+1}) = \beta_0(\tau) + \beta_d^+(\tau)RS_t^+ + \beta_d^-(\tau)RS_t^- + \gamma(\tau)RV_t I(r_t < 0) + \beta_w(\tau)RV_{w,t} + \beta_m(\tau)RV_{m,t}. \tag{3.11}$$

Model 7: The third model, presented by HARQ-RV-SJ-I, contains a signed jump variance and BV_t as follows:

$$q_\tau(RV_{t+1}) = \beta_0(\tau) + \beta_{j,d}(\tau)SJ_t + \beta_{bv,d}(\tau)BV_t + \beta_w(\tau)RV_{w,t} + \beta_m(\tau)RV_{m,t}, \tag{3.12}$$

where $SJ_t = RS_t^+ - RS_t^-$ is the signed jump variation.

Model 8:. The fourth model, to assess the different impacts between positive jump variation and negative jump variation, named by HARQ-RV-SJ-II, is designed by

$$q_\tau(RV_t) = \beta_0(\tau) + \beta_{j,d}^+(\tau)SJ_t^+ + \beta_{j,d}^-(\tau)SJ_t^- + \beta_{bv,d}(\tau)BV_t + \beta_w(\tau)RV_{w,t} + \beta_m(\tau)RV_{m,t}, \tag{3.13}$$

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

Model 9: By following the idea in Corsi and Renò (2012) for the volatility forecasting, the heterogeneous structure to the standard leverage effect is extended by including lagged negative returns at different frequencies as explanatory variables to forecast volatility under the quantile setting, which is called the leverage heterogeneous autoregressive with continuous volatility and jumps quantile model (LHARQ-CJ), given by

$$\begin{aligned} q_\tau(RV_{t+1}) =& \beta_0(\tau) + \beta_{c,d}(\tau)TC_t + \beta_{c,w}(\tau)TC_{w,t} + \beta_{c,m}(\tau)TC_{m,t} + \beta_{j,d}(\tau)TJ_t + \beta_{j,w}(\tau)TJ_{w,t} \\ &+ \beta_{j,m}(\tau)TJ_{m,t} + \gamma_d(\tau)r_t I(r_t < 0) + \gamma_w(\tau)r_{w,t} I(r_{w,t} < 0) + \gamma_m(\tau)r_{m,t} I(r_{m,t} < 0), \end{aligned} \tag{3.14}$$

where TC_t and TC_t is defined in (3.9), and $r_{w,t}$ is defined in the same fashion as $RV_{w,t}$ and similar to $r_{m,t}$.

Model 10: Motivated by the approaches in Audrino and Hu (2016) for the volatility forecasting, to disentangle continuous variation and jump variation via complete decomposition

of signed continuous and jump variation, we consider the first model, denoted by LHARQ-JSV-I,

$$\begin{aligned} q_\tau(RV_{t+1}) =& \beta_0(\tau) + \beta_d^+(\tau)CSV_t^+ + \beta_d^-(\tau)CSV_t^- + \beta_{j,d}^+(\tau)JSV^+ \\ &+ \beta_{j,d}^-(\tau)JSV_t^- + \theta(\tau)VolJ_t + \delta(\tau)cRet_t^- + \delta_w(\tau)cRet_{w,t}^- \\ &+ \beta_w(\tau)RV_{w,t} + \beta_m(\tau)RV_{m,t}, \end{aligned} \tag{3.15}$$

where JSV_t^+ and JSV_t^- are the jump semi-variances, $CSV_t^+ = RS_t^+ - JSV_t^+$ and $CSV_t^- = RS_t^- - JSV_t^-$ are the continuous semi-variances (good and bad volatilities), $VolJ_t$ is the jump in volatility, and $cRet_t^-$ is the jump-adjusted return.

Model 11: To simplify the above model and improve the out-of-sample results, by following Audrino and Hu (2016) with focus on the most relevant effects for volatility forecasting: downside risk, leverage effect and the HAR structure capturing long memory, we consider the following model, denoted as LHARQ-JSV-II

$$\begin{aligned} q_\tau(RV_{t+1}) =& \beta_0(\tau) + \beta_d^+(\tau)CSV_t^+ + \beta_d^-(\tau)CSV_t^- + \delta(\tau)cRet_t^- \\ &+ \varphi(\tau)jRet_t + \beta_w(\tau)RV_{w,t} + \beta_m(\tau)RV_{m,t} \end{aligned} \tag{3.16}$$

where $jRet_t$ is the jump size.

Finally, note that for the detail definitions and calculations of variables in Models 1 – 11 defined in (3.6) – (3.16), the reader is referred to the corresponding references as mentioned above. For simplicity, these eleven models are denoted as HARQi, where i ($1 \leq i \leq 11$) is the corresponding number of each model. Besides, for brevity, it is called the realized measures with subscript t,w and m as daily, weekly and monthly realized measures, e.g., RV_t is called daily RV.

3.3.2 Results from Quantile Models

Tables 2 and 3 show the results of the eleven HARQ models over the whole sample period along with p-values of the t-statistics in brackets. For illustration and brevity, this paper presents only the results for five quantiles (namely, $\tau = 0.1$, 0.25, 0.5, 0.75, and 0.9). From these two tables, one can observe the following results.

First, all estimators of the parameters of RV_t, $RV_{w,t}$ and $RV_{m,t}$ (namely β_d, β_w and β_m)

are significantly positive at the 1% significance level in all models at each quantile, except three estimators at quantile 0.9, which provides a strong evidence to support properties of the strong persistence and long memory in quantiles of RV. Here, two observations can be made. On the one hand, there is a clear increasing trend: the daily, weekly and monthly RV have larger impacts on higher quantiles of future volatility. This result not only is consistent with common sense that volatility is clustered and has long memory but also indicates that higher RV could induce a fatter right tail in the future with stronger persistence in high-level quantiles, namely, increase the probability of producing a much larger value. This result is also consistent with the findings of volatility-of-volatility effects in Corsi et al. (2008) and Bollerslev et al. (2009). On the other hand, the daily RV has larger coefficient than the weekly and the monthly RV. However, this is not true for weekly RV and monthly RV. Although the weekly RV has larger impact than monthly RV at higher quantiles, the impacts at lower quantiles are just opposite. Therefore, one can conclude that the long memory in high-level quantiles declines faster than that in low-level quantiles.

Second, the coefficients of the continuous component (see $\beta_{c,d}$, $\beta_{c,w}$, $\beta_{c,m}$) have similar changing patterns with lags of RV for different quantiles. While, the impacts of the jumps are more complicated and mixed with no fixed impact patterns on different quantiles for different measures of the jumps. They can even be negative, which increases the difficulty of volatility forecasting. However, from HARQ10, one can see that bad jumps have positive effects ($\beta^-_{j,d}$) on future quantiles of RV while good jumps are opposite. Both components of jumps have larger absolute effects on higher-level quantiles. This means that bad jumps can amplify the future outcomes and right tails of RV while good jumps downsize and stable the outcomes of RV.

Third, we find strong support that bad volatility (β^-_d) has larger effect than good volatility (β^+_d) with much larger coefficient. Similarly, bad volatility has an increasing effect as the quantiles increase. However, good volatility has much weaker smooth flat effects on different quantiles and can even have a negative effect at the high quantile of $\tau = 0.9$. Thus, bad volatility leads to higher volatility and fatter tails.

Last, the leverage effects of returns (γ) are significant at all horizons for different quantiles. Notably, jump-adjusted daily returns have less impacts than total returns on volatility.

Furthermore, the leverage effect is also much larger at higher quantile.

In summary, the stylized features of RV are stronger at high-level quantiles, indicating stronger predictive powers for high-level RV quantiles.

Table 2: Estimations of realized volatility quantile models for the whole sample

	HARQ1					HARQ2					HARQ3					HARQ4					HARQ5				
	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9
β_0	**-0.002**	**0.001**	0.018	0.059	0.119	-0.003	**0.002**	0.018	0.057	0.118	**-0.002**	**-0.001**	0.015	0.048	0.116	**-0.001**	**0.002**	0.022	0.065	0.131	-0.006	**-0.002**	0.013	0.064	0.133
	(0.467)	(0.657)	(0.000)	(0.000)	(0.000)	(0.253)	(0.519)	(0.000)	(0.000)	(0.000)	(0.352)	(0.784)	(0.000)	(0.000)	(0.000)	(0.785)	(0.332)	(0.000)	(0.000)	(0.000)	(0.039)	(0.552)	(0.000)	(0.000)	(0.000)
β_d	0.270	0.370	0.473	0.657	0.765	0.270	0.342	0.472	0.671	0.823															
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)															
β_w	0.025	0.074	0.159	0.331	0.647	0.026	0.074	0.160	0.372	0.648											0.032	0.086	0.149	0.354	0.672
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)											(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β_m	0.148	0.164	0.162	0.090	**-0.023**	0.146	0.164	0.164	0.075	**-0.031**											0.148	0.148	0.173	0.125	**0.018**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.141)	(0.000)	(0.000)	(0.000)	(0.000)	(0.061)											(0.000)	(0.000)	(0.000)	(0.000)	(0.242)
$\beta_{c,d}$											0.270	0.374	0.468	0.673	0.781	0.267	0.364	0.492	0.719	0.861					
											(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
$\beta_{c,w}$											0.025	0.074	0.165	0.318	0.603	0.030	0.087	0.161	0.284	0.585					
											(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
$\beta_{c,m}$											0.148	0.163	0.152	0.090	**-0.013**	0.150	0.159	0.144	0.101	-0.031					
											(0.000)	(0.000)	(0.000)	(0.000)	(0.427)	(0.000)	(0.000)	(0.000)	(0.000)	(0.046)					
$\beta_{j,d}$						0.047	0.230	-0.039	-0.313	-0.733	-0.091	0.142	**0.057**	**0.074**	**-0.179**	0.125	0.402	0.457	0.350	0.347					
						(0.008)	(0.000)	(0.104)	(0.000)	(0.000)	(0.010)	(0.001)	(0.283)	(0.520)	(0.438)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
$\beta_{j,w}$											0.435	0.659	1.230	1.610	1.142										
											(0.000)	(0.000)	(0.000)	(0.000)	(0.052)										
$\beta_{j,m}$											**0.006**	**0.071**	1.023	1.712	**1.811**										
											(0.973)	(0.731)	(0.000)	(0.003)	(0.112)										
β_d^+																					0.115	0.158	0.139	-0.081	-0.396
																					(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β_d^-																					0.432	0.622	0.862	1.217	1.798
																					(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Note: The second row is the quantile points. The values in the brackets are the p-values of the t-statistics. The insignificant results are marked in bold.

Table 3: Estimations of realized volatility quantile models for the whole sample

	HARQ6					HARQ7					HARQ8				
	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9
β_0	-0.006	**-0.001**	0.010	0.070	0.130	**-0.003**	0.002	0.015	0.068	0.136	**-0.002**	**-0.001**	0.012	0.071	0.121
	(0.018)	(0.564)	(0.001)	(0.000)	(0.000)	(0.237)	(0.427)	(0.000)	(0.000)	(0.000)	(0.492)	(0.744)	(0.000)	(0.000)	(0.000)
β_d															
β_w	0.039	0.086	0.154	0.333	0.641	0.029	0.102	0.217	0.429	0.703	0.035	0.112	0.207	0.390	0.774
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β_m	0.143	0.151	0.162	0.122	0.064	0.158	0.165	0.169	0.118	**0.013**	0.146	0.172	0.173	0.126	**0.004**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.288)	(0.000)	(0.000)	(0.000)	(0.000)	(0.792)
$\beta_{c,d}$															
$\beta_{c,w}$															
$\beta_{c,m}$															
$\beta_{j,d}$						-0.207	-0.212	-0.385	-0.587	-0.976					
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
$\beta_{j,w}$															
$\beta_{j,m}$															
β_d^+	0.065	0.069	0.052	0.120	0.198										
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)										
β_d^-	0.165	0.256	0.193	0.081	-0.198										
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)										
γ	0.315	0.439	0.799	0.936	1.378										
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)										
γ_d															
γ_w															
γ_m															
$\beta_{bv,d}$						0.274	0.356	0.468	0.519	0.714	0.223	0.288	0.400	0.446	0.482
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\beta_{j,d}^+$											0.093	**-0.008**	-0.092	-0.194	-0.273
											(0.000)	(0.516)	(0.000)	(0.000)	(0.000)
$\beta_{j,d}^-$											-0.448	-0.585	-0.826	-1.114	-2.104
											(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

	HARQ9					HARQ10					HARQ11				
	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9
β_0	**0.002**	**0.000**	**0.001**	**0.009**	0.050	-0.010	**0.001**	**0.004**	0.050	0.110	-0.009	**0.004**	0.011	0.065	0.131
	(0.682)	(0.946)	(0.841)	(0.295)	(0.010)	(0.002)	(0.826)	(0.325)	(0.000)	(0.000)	(0.001)	(0.121)	(0.002)	(0.000)	(0.000)
β_d															
β_w						-0.011	0.030	0.137	0.171	0.347	-0.016	0.034	0.124	0.210	0.537
						(0.026)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β_m						0.144	0.148	0.144	0.130	0.063	0.143	0.140	0.144	0.090	**-0.006**
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.668)
$\beta_{c,d}$	0.247	0.310	0.441	0.547	0.621										
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)										
$\beta_{c,w}$	0.030	0.120	0.164	0.267	0.481										
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)										
$\beta_{c,m}$	0.205	0.167	0.160	0.110	**0.007**										
	(0.000)	(0.000)	(0.000)	(0.000)	(0.682)										
$\beta_{j,d}$	0.159	0.328	0.455	0.280	0.136										
	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)										
$\beta_{j,w}$	**-0.035**	-0.246	-0.358	**0.081**	0.431										
	(0.193)	(0.000)	(0.000)	(0.074)	(0.000)										
$\beta_{j,m}$	0.214	0.850	1.176	1.030	**0.086**										
	(0.000)	(0.000)	(0.000)	(0.000)	(0.694)										
β_d^+						0.171	0.154	0.167	0.183	-0.191	0.187	0.131	0.244	0.381	-0.171
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β_d^-						0.572	0.727	0.934	1.162	1.979	0.573	0.782	0.940	1.161	1.922
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
γ															
γ_d	-0.070	-0.100	-0.134	-0.220	-0.406										
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)										
γ_w	-0.058	-0.025	-0.119	-0.285	-0.362										
	(0.000)	(0.042)	(0.000)	(0.000)	(0.000)										
γ_m	0.199	0.122	0.084	**-0.038**	-0.273										
	(0.000)	(0.000)	(0.001)	(0.381)	(0.009)										
$\beta_{bv,d}$															
$\beta_{j,d}^+$						-0.033	-0.190	-0.102	-0.142	-0.077					
						(0.038)	(0.000)	(0.000)	(0.000)	(0.285)					
$\beta_{j,d}^-$						0.105	0.255	0.558	0.585	0.394					
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
θ						0.128	0.361	0.101	-0.076	-0.236					
						(0.000)	(0.000)	(0.000)	(0.001)	(0.000)					
δ						-0.016	-0.017	-0.044	-0.050	-0.042	**-0.009**	-0.018	-0.073	-0.109	-0.167
						(0.042)	(0.019)	(0.000)	(0.004)	(0.229)	(0.144)	(0.009)	(0.000)	(0.000)	(0.000)
δ_w						**0.004**	-0.030	-0.096	-0.292	-0.423					
						(0.781)	(0.022)	(0.000)	(0.000)	(0.000)					
φ											**-0.001**	-0.020	-0.063	-0.114	-0.110
											(0.906)	(0.046)	(0.000)	(0.000)	(0.016)

Note: The second row is the quantile points. The values in the brackets are the p-values of the t-statistics. The insignificant results are marked in bold.

3.4 RV Forecasting Performance of DQMA

After the in-sample analyses, we move to evaluate the out-of-sample forecasting performance, which might be more important for practitioners. First, the evaluation criteria are provided for forecasting performance and then followed by some empirical results. Additionally, as argued in Wang et al. (2016), DMA-TVP model has a good performance for RV mean prediction, which outperforms eighteen other strategies, so that DMA-TVP is selected as a competing strategy for our DQMA models. Another popular competing model used to capture the long-memory features of RV is the ARFIMA(p, d, q) model

$$\phi(B)(1-B)^d \log(RV_t) = \psi(B)u_t,$$

where B denotes the lag operator, $\phi(B)$ is the pth order of polynomial of B, $\psi(B)$ is the qth order of polynomial of B, $|d| < 1/2$, and u_t is a white noise series. This model is often regarded by researchers as a benchmark for modeling volatility; see, for example, Andersen et al. (2003) and Corsi (2009). Moreover, this model was also used in Žikeš and Baruník (2016) as a competing model with their three HARQ models.

3.4.1 Evaluation Criteria

Several evaluation methods exist for quantile (or VaR) forecast; see, for examples, Kupiec (1995), Christoffersen (1998), Engle and Manganelli (2004), McAleer and Da Veiga (2008) and the references therein. In particular, Berkowitz et al. (2011) provided an integrated and unifying framework for assessing the accuracy for VaR forecasts. The accuracy of a set of quantile forecasts can be assessed by viewing them as one-sided interval forecasts. A hit variable is usually defined as follows:

$$Hit_{t+1} = I(RV_{t+1} \leq q_{t+1,\tau}),$$

which is a binary variable, equal to one if the conditional quantile is violated and zero otherwise. For the assessment of symmetric interval forecasting (two-sided), another hit variable, according to Christoffersen (1998), is defined by

$$Hit^{ts}_{t+1} = I(q_{t+1,\tau} \leq RV_{t+1} \leq q_{t+1,1-\tau}).$$

If the conditional quantiles are correctly forecasted, hits are sequenced independently with expected probability of τ (or $1-2\tau$ in interval forecasting). However, for asymmetric interval forecasting, a three-state Markov transition matrix is needed; see Christoffersen (1998) for details. This hit variable can then be used to construct tests or evaluations for quantile forecasting. In this paper, five evaluation methods are used. The first is the UC test of Kupiec (1995), the second is the CC test of Christoffersen (1998), the third is the DQ test of Engle and Manganelli (2004), for which 4 lags are chosen as suggested, the fourth is AE, the actual-over-expected ratio, and the last is the coverage rate of interval forecasting. Furthermore, to evaluate the quantile forecasting accuracy, following Koenker and Machado (1999), we suggest using the check loss functions and the goodness of fit of the quantiles, defined in (2.11).

3.4.2 Results

Clearly, the in-sample predictive relationships are not constant but time-varying, which can be demonstrated if samples are separated into different intervals and then estimate models in these intervals. By doing so, one can obtain different estimators for the same model in different sample periods. In fact, Wang et al. (2016) showed that models with time-varying coefficients are superior to those with constant coefficients. To investigate the out-of-sample predictability and mitigate the effects of time-variation, rolling windows are used to estimate models and realized volatilities are recursively predicted, thereby producing a sequence of pseudo predictors. While the rolling estimator may appear similar to a parametric estimator, it is a local constant estimator and, thus, a nonparametric time-varying parameter $\beta_{t,\tau}$, where the estimation window size plays the role of the bandwidth; see Inoue et al. (2017) for details. Hence, the coefficients of all models are in fact time-varying parameters $\beta_{t,\tau,k}$ in this section. In terms of how to choose the rolling window size, it is a challenging issue in quantile regression setting but it is still open. Therefore, after conducting many trials, 500 samples are chosen as the rolling window size, which is the same number used in Žikeš and Baruník (2016). For the prior parameters of DMA, the forgetting factor α_0 in (2.3) is set to be 0.99 according to the suggestions of Raftery et al. (2010) and Koop and Korobilis (2012). In estimating the recursive moment estimator in DMA, we follow the same procedures as in Koop and

Korobilis (2012), by using the exponentially weighted moving average (EWMA) and the same values for the related parameters for daily data. All the initial variances are equal to 10, a value that is commonly used in volatility modeling in DMA. For ARFIMA model, ARFIMA$(1, d, 0)$ is used as suggested in Žikeš and Baruník (2016). As for the sequence quantile points τ's selection in estimating (2.5), we choose the range from 0.05 to 0.95 in intervals of 0.05 and add two points (0.01 and 0.99) at either end. Clearly, other sequences can be chosen if desired. In fact, the same experiment is repeated with the range from 0.1 to 0.9 in intervals of 0.1 and similar empirical results can be obtained.

Table 4: Goodness of fit R^1_τ, where the rolling window size is 500

τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
HARQ1	0.1922	0.2401	0.2766	0.3175	0.3539	0.3865	0.4315	0.4826	0.5511
HARQ2	0.1918	0.2300	0.2720	0.3145	0.3548	0.3850	0.4328	0.4770	0.5531
HARQ3	0.1764	0.2251	0.2706	0.3142	0.3490	0.3835	0.4286	0.4827	0.5450
HARQ4	0.1815	0.2297	0.2722	0.3134	0.3509	0.3840	0.4332	0.4803	0.5567
HARQ5	0.1882	0.2418	0.2868	0.3270	0.3632	0.4077	0.4443	0.5002	0.5543
HARQ6	0.1866	0.2449	0.2911	0.3282	0.3644	0.4047	0.4502	0.5002	0.5565
HARQ7	0.1869	0.2391	0.2855	0.3220	0.3623	0.3991	0.4416	0.4957	0.5586
HARQ8	0.1917	0.2388	0.2947	0.3322	0.3683	0.4075	0.4475	0.5008	0.5537
HARQ9	0.1787	0.2414	0.2922	0.3218	0.3568	0.3981	0.4360	0.4702	0.5476
HARQ10	0.1488	0.2234	0.2794	0.3262	0.3703	0.3355	0.4074	0.4679	0.5368
HARQ11	0.2003	**0.2531**	**0.2980**	**0.3390**	**0.3746**	**0.4158**	**0.4593**	**0.5086**	**0.5738**
DQMA-I	0.1943	0.2482	0.2959	0.3374	0.3692	0.3890	0.4351	0.4793	0.5464
DQMA-II	0.2014	0.2494	0.2948	0.3357	0.3733	0.4049	0.4513	0.5002	0.5633
DQMA-Eq	0.2009	0.2494	0.2942	0.3335	0.3719	0.4041	0.4495	0.4993	0.5621
DQMA-Gods	0.2008	0.2493	0.2941	0.3334	0.3717	0.4044	0.4499	0.4997	0.5625
DQMA-MSE	**0.2019**	0.2487	0.2943	0.3341	0.3726	0.4007	0.4477	0.4983	0.5610

Note: The first row is the quantile sequence points. We present only nine points for illustration. The number in bold is the largest in the specific quantile. Similarly, the number in red is the smallest.

First, we consider the goodness of fit of individual HARQ and DQMA models in quantile forecasting, as shown in Table 4. Obviously, the DQMA models never have the smallest goodness of fit at each quantile. In fact, the values of DQMA models are always being or close to the best one in different quantiles. Furthermore, the goodness of fit for the DQMA models is more stable than individual HARQ. This provides a strong evidence to conclude that the combination method is a good choice to address the risk of model uncertainty. In fact, if one chooses a worse variable or artificially contaminate one variable in an individual model such that the intervened model induces a worse forecasting performance, one can obtain much more significant and clear superiority of out-of-sample performance of the DQMA methods

(we do not present this manually interfering experiment in this paper). HARQ11 appears to be the best individual model in terms of fit because it performs the best in eight scenarios of quantiles. Similarly, HARQ3 and HARQ10 perform worse. Table 4 also indicates that the goodness of fit is better at high quantiles than low quantiles. This means that quantile regression is more powerful and suitable for forecasting rare extremely large realizations of volatility. This ability is very important for investors exposing to risks. In an economic sense, forecasting extreme increasing RV events is much more important for risk measurement.

Table 5 reports the AE ratio for all models from quantiles 0.1 to 0.9. The numbers in red, which indicate values far from one, all come from ARFIMA and DMA, as expected. Again, although the DQMA models do not always perform the best, they are never the worst. Table 5 provides evidences that DMA and ARFIMA tend to underestimate the low-level quantiles and overestimate high-level quantiles, resulting in low AE ratios for low quantiles and high AE ratios for high quantiles, as well as large negative deviations for upper interval forecasting (see Table 7 for details).

Table 5: AE quantile exceedance in out-of-sample

τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
ARFIMA	0.0000	0.1368	0.7907	1.3191	1.4607	1.4285	1.3265	1.2082	1.0998
DMA	0.4866	0.6918	0.9427	1.1328	1.2164	1.2249	1.1875	1.1138	1.0368
HARQ1	1.1809	1.1252	1.0813	1.0403	1.0177	1.0078	0.9978	1.0016	0.9957
HARQ2	1.1556	1.1100	1.0897	1.0428	1.0208	1.0128	0.9963	1.0016	0.9979
HARQ3	1.1505	1.1353	1.0779	1.0378	**1.0117**	1.0052	0.9934	0.9953	0.9923
HARQ4	1.2215	1.1379	1.0982	1.0454	1.0258	1.0078	**1.0007**	1.0010	0.9957
HARQ5	1.1860	1.1024	1.0610	1.0251	1.0309	1.0120	1.0064	**0.9997**	0.9951
HARQ6	1.1809	1.0897	1.0542	1.0289	1.0238	1.0095	0.9956	1.0023	0.9945
HARQ7	1.1809	1.1024	1.0880	1.0454	1.0279	1.0069	0.9963	1.0067	0.9974
HARQ8	1.1708	1.0973	1.0846	1.0390	1.0177	**1.0010**	0.9956	0.9953	0.9957
HARQ9	1.2316	1.1531	1.0542	**1.0061**	1.0208	1.0137	0.9898	0.9807	0.9810
HARQ10	1.1404	1.1125	1.0542	1.0542	1.0319	1.0103	1.0093	1.0029	0.9957
HARQ11	**1.1303**	**1.0897**	**1.0306**	1.0352	1.0228	0.9909	1.0028	**0.9997**	0.9945
DQMA-I	1.1607	1.1353	1.0644	1.0416	1.0380	1.0272	1.0130	0.9991	0.9928
DQMA-II	1.1860	1.1201	1.0610	1.0264	1.0248	1.0188	1.0108	1.0061	**1.0007**
DQMA-Eq	1.1911	1.1201	1.0745	1.0327	1.0279	1.0179	1.0115	1.0105	1.0013
DQMA-Gods	1.1961	1.1201	1.0745	1.0314	1.0269	1.0188	1.0122	1.0112	**1.0007**
DQMA-MSE	1.1860	1.1176	1.0711	1.0289	1.0279	1.0204	1.0115	1.0105	**1.0007**

Note: The numbers in bold are those closest to one in each column, whereas the numbers in red are those farthest from one.

Table 6 shows the p-values of the UC test, CC test and DQ test for the quantile forecasting performance. First, Table 6 suggests that for forecasting the low-level quantiles of RV, all the models and methods are somewhat misspecified and inadequate. However,

Table 6: p-Values of the UC test, CC test and DQ test for out-of-sample performance

τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
					UC test p-Value				
ARFIMA	0.0000	0.0000	0.0940	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
DMA	0.0000	0.0000	0.0940	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
HARQ1	0.0090	0.0062	0.0190	0.1448	0.4307	0.6723	0.8790	0.8836	0.5656
HARQ2	0.0242	0.0159	0.0097	0.1212	0.3560	0.4844	0.8023	0.8836	0.7819
HARQ3	0.0291	0.0031	0.0245	0.1717	0.6046	0.7756	0.6553	0.6778	0.3088
HARQ4	0.0015	0.0026	0.0047	0.1007	0.2509	0.6723	0.9647	0.9282	0.5656
HARQ5	0.0073	0.0247	0.0778	0.3635	0.1696	0.5136	0.6615	0.9820	0.5165
HARQ6	0.0090	0.0487	0.1166	0.2955	0.2900	0.6065	0.7646	0.8393	0.4698
HARQ7	0.0090	0.0247	0.0111	0.1007	0.2156	0.7061	0.8023	0.5494	0.7252
HARQ8	0.0135	0.0327	0.0146	0.1578	0.4307	0.9560	0.7646	0.6778	0.5656
HARQ9	0.0009	0.0008	0.1166	0.8255	0.3560	0.4561	0.4895	0.0901	0.0135
HARQ10	0.0416	0.0137	0.1166	0.0498	0.1561	0.5747	0.5254	0.7954	0.5656
HARQ11	0.0583	0.0487	0.3753	0.2023	0.3110	0.6199	0.8480	0.9820	0.4698
DQMA-I	0.0200	0.0031	0.0628	0.1326	0.0913	0.1380	0.3778	0.9372	0.3451
DQMA-II	0.0073	0.0085	0.0778	0.3398	0.2699	0.3069	0.4631	0.5878	0.9222
DQMA-Eq	0.0059	0.0085	0.0314	0.2366	0.2156	0.3292	0.4335	0.3476	0.8627
DQMA-Gods	0.0047	0.0085	0.0314	0.2552	0.2328	0.3069	0.4051	0.3191	0.9222
DQMA-MSE	0.0073	0.0100	0.0399	0.2955	0.2156	0.2653	0.4335	0.3476	0.9222
					CC test p-Value				
ARFIMA	0.0000	0.0000	0.0940	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
DMA	0.0000	0.0000	0.0045	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
HARQ1	0.0019	0.0001	0.0000	0.0072	0.4144	0.0196	0.9228	0.7670	0.8057
HARQ2	0.0037	0.0001	0.0000	0.0019	0.3712	0.3601	0.6582	0.7742	0.8989
HARQ3	0.0007	0.0006	0.0000	0.0009	0.1240	0.0295	0.7473	0.7844	0.5636
HARQ4	0.0006	0.0001	0.0001	0.0098	0.2691	0.0411	0.9735	0.8714	0.8363
HARQ5	0.0027	0.0216	0.0479	0.1185	0.3581	0.5441	0.1828	0.7137	0.8039
HARQ6	0.0126	0.1399	0.1487	0.1125	0.4328	0.8723	0.9538	0.7280	0.7357
HARQ7	0.0001	0.0174	0.0001	0.0191	0.0824	0.1164	0.0463	0.5640	0.8938
HARQ8	0.0001	0.0722	0.0005	0.0569	0.2993	0.8467	0.3340	0.7844	0.8454
HARQ9	0.0001	0.0030	0.0843	0.1972	0.4068	0.6777	0.5384	0.1795	0.0161
HARQ10	0.0241	0.0476	0.2846	0.0106	0.3478	0.6356	0.7750	0.7636	0.6068
HARQ11	0.1523	0.1360	0.5295	0.3045	0.5653	0.3968	0.5915	0.6866	0.5066
DQMA-I	0.0058	0.0061	0.0105	0.1199	0.2402	0.2230	0.5022	0.8858	0.6332
DQMA-II	0.0041	0.0089	0.0036	0.0264	0.4533	0.5921	0.7578	0.7211	0.3541
DQMA-Eq	0.0038	0.0089	0.0014	0.0886	0.4638	0.5982	0.7334	0.4675	0.5276
DQMA-Gods	0.0034	0.0089	0.0014	0.0961	0.4905	0.5745	0.6996	0.4616	0.5046
DQMA-MSE	0.0041	0.0114	0.0011	0.0654	0.4258	0.5257	0.7334	0.5202	0.5046
					DQ test p-Value				
ARFIMA	0.0000	0.0000	0.0940	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
DMA	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
HARQ1	0.0000	0.0000	0.0000	0.0001	0.1237	0.0021	0.0070	0.1694	0.3688
HARQ2	0.0000	0.0000	0.0000	0.0001	0.1245	0.0052	0.0196	0.6367	0.5370
HARQ3	0.0000	0.0000	0.0000	0.0001	0.0503	0.0001	0.0919	0.2119	0.8952
HARQ4	0.0000	0.0000	0.0000	0.0000	0.0092	0.0023	0.0768	0.4795	0.9077
HARQ5	0.0000	0.0000	0.0000	0.0002	0.0526	0.0224	0.0748	0.1789	0.2324
HARQ6	0.0000	0.0000	0.0000	0.0182	0.1766	0.2635	0.4922	0.3822	0.2095
HARQ7	0.0000	0.0000	0.0000	0.0000	0.0024	0.0156	0.0169	0.5026	0.3027
HARQ8	0.0000	0.0000	0.0000	0.0001	0.0030	0.0085	0.0342	0.2998	0.8861
HARQ9	0.0000	0.0000	0.0001	0.0045	0.0107	0.0191	0.0198	0.1218	0.0482
HARQ10	0.0000	0.0000	0.0037	0.0002	0.1309	0.2108	0.3473	0.6547	0.7619
HARQ11	0.0003	0.0001	0.0042	0.0008	0.2470	0.5015	0.6911	0.4652	0.8985
DQMA-I	0.0000	0.0000	0.0000	0.0911	0.0675	0.0027	0.0257	0.2326	0.3091
DQMA-II	0.0000	0.0000	0.0000	0.0007	0.1277	0.0549	0.2135	0.3911	0.8124
DQMA-Eq	0.0000	0.0000	0.0000	0.0009	0.0887	0.0130	0.1187	0.2796	0.7714
DQMA-Gods	0.0000	0.0000	0.0000	0.0009	0.0870	0.0132	0.1442	0.3164	0.7912
DQMA-MSE	0.0000	0.0000	0.0000	0.0007	0.1153	0.0188	0.1209	0.2836	0.7467

Note: The numbers in red are significantly different from the null hypothesis at the 1% significance level, which indicates misspecification of the quantile model.

Table 7: Evaluation of the interval forecasting in the out-of-sample

τ interval	(0.05,0.95)			(0.1,0.9)			(0.25,0.75)			(0.45,0.95)		
	length	p-0.9	Lcc	length	p-0.8	Lcc	length	p-0.5	Lcc	length	p-0.5	Lcc
ARFIMA	0.7945	0.0944	0.0000	0.6190	0.1899	0.0000	0.3258	0.3510	0.0000	**0.4276**	-0.1488	0.0000
DMA	1.4041	0.0311	0.0000	1.0940	0.0844	0.0000	0.5758	0.1569	0.0000	0.7557	-0.0819	0.0000
HARQ1	0.7780	-0.0191	0.0001	0.5431	-0.0220	0.0051	0.2740	-0.0215	0.1057	0.6105	-0.0175	0.1599
HARQ2	0.7833	-0.0110	0.0001	0.5443	**-0.0174**	0.0240	0.2790	-0.0190	0.0663	0.6129	-0.0139	0.1134
HARQ3	0.8055	-0.0196	0.0000	0.5476	-0.0220	0.0021	0.2760	-0.0261	0.0482	0.6334	**-0.0139**	0.0824
HARQ4	0.7800	-0.0191	0.0000	0.5470	-0.0261	0.0006	0.2823	-0.0215	0.0455	0.6122	-0.0246	0.0075
HARQ5	0.7654	-0.0196	0.0002	0.5476	-0.0230	0.0016	0.2615	-0.0195	0.0246	0.5943	-0.0180	0.0705
HARQ6	0.7637	-0.0201	0.0000	0.5523	-0.0230	0.0050	0.2587	-0.0180	0.0705	0.5944	-0.0170	0.1372
HARQ7	0.7764	-0.0201	0.0000	0.5491	-0.0205	0.0013	0.2664	-0.0266	0.0385	0.5986	-0.0180	0.0039
HARQ8	0.7578	-0.0211	0.0001	0.5488	-0.0210	0.0006	0.2583	-0.0190	0.1045	0.5851	-0.0200	0.0328
HARQ9	**0.7554**	-0.0318	0.0000	**0.5358**	-0.0397	0.0000	**0.2583**	-0.0388	0.0010	0.5922	-0.0150	0.0432
HARQ10	0.7828	-0.0176	0.0020	0.5733	-0.0179	0.0080	0.2889	-0.0139	0.1721	0.6161	-0.0205	0.0305
HARQ11	0.7807	-0.0171	0.0052	0.5442	-0.0179	0.1000	0.2700	-0.0114	0.3256	0.6116	-0.0215	0.0422
DQMA-I	0.7838	-0.0191	0.0025	0.5564	-0.0225	0.0066	0.2704	-0.0139	0.2309	0.6154	-0.0236	0.0625
DQMA-II	0.7720	-0.0105	0.0058	0.5522	-0.0179	0.0306	0.2693	-0.0079	0.7115	0.6016	-0.0185	0.1535
DQMA-Eq	0.7754	-0.0090	0.0049	0.5485	-0.0179	0.0381	0.2703	-0.0084	0.6265	0.6056	-0.0150	0.2177
DQMA-Gods	0.7746	-0.0090	0.0049	0.5475	-0.0190	0.0299	0.2701	-0.0084	0.6265	0.6048	-0.0150	0.1746
DQMA-MSE	0.7758	**-0.0080**	0.0042	0.5496	-0.0179	0.0381	0.2711	**-0.0073**	0.7330	0.6060	-0.0165	0.1819

Note: The numbers in bold are the shortest in the length columns (the length of $q_{\tau_{i+1}} - q_{\tau_i}$) and are the closest to the expected percentage in the p-x columns (the length of the τ interval, $\tau_{i+1} - \tau_i$). The numbers in red are the opposite. The red values in Lcc (CC test of Christoffersen (1998)) are those less than the 1% confidence level for which the null hypothesis is rejected.

except ARFIMA and DMA, almost all the models accept the null hypothesis of correct dynamic specification for high-level quantiles of RV. This can be explained by our finding in the above subsection that the high-level quantiles of RV have stronger predictabilities and stronger stylized features. Second, all the tests for ARFIMA and DMA models reject the null hypothesis, demonstrating their inabilities to forecast or model the quantile dynamics of RV. Third, the DQMA models are more stable than the individual quantile models, especially in the DQ test. Take the DQ test at $\tau = 0.5$ for example, there are two individual HARQ models that reject the null hypothesis but no DQMA models. Moreover, even though five individual HARQ models reject the null hypothesis of the DQ test at $\tau = 0.6$, only one of DQMA (DQMA-I) model does.

Table 7 shows the evaluation of interval forecasting for the HARQ, DQMA, ARFIMA and DMA models. For simplicity, we present only the CC tests of Christoffersen (1998) for interval forecasting (the L_{cc} column). Here, one can see again the stabilities of DQMA models contrasted to individual quantile models and the advantages of quantile models compared with mean models. First, the average longest interval lengths are produced by DMA mod-

els(the length columns). As well, the largest deviations from the expected percentages come from ARFIMA(the p-x columns). The performances of ARFIMA and DMA are especially bad for upper asymmetric interval forecasting. Take the interval (0.45,0.95) for example; the deviation percentage can reach 14.88% for ARFIM. Clearly, the quantile models are found to be more suitable than mean models for RV interval forecasting. Second, although the lengths of forecasting intervals obtained from HARQ and DQMA models are relatively equal, especially for $\tau \subset (0.05, 0.95)$, the deviation percentages of DQMA are smaller than those of HARQ overall. More importantly, it can be evidenced from L_{cc} that the DQMA models are more stable than HARQ, e.g., in the interval $\tau \subset (0.1, 0.9)$, nine individual HARQ models reject the null hypothesis of the CC tests, but only one DQMA model (DQMA-I) rejects the null hypothesis. Furthermore, two individual models reject the null hypothesis for the interval $(0.45, 0.95)$, but no DQMA models do. These results clearly indicate that DQMA can not only achieve better performance but also reduce the risk of model uncertainty.

Figure 2 shows the dynamics of the S&P 500 index daily price and combination weight evolution for the different combining strategies. From Figure 2, it can be easily seen that different combination strategies have dramatically different weights. The weights with the greatest changes are those of the DQMA-I strategy, followed by those of the DMA method. The other strategies are smoother and more stable. The dynamics of DQMA-MES are almost horizontal with nearly constant weights. In addition, the dynamic evolutions of the weights may provide a picture of the structural changes of RV. See, the second row (DQMA-II) and the last row (DQMA-Gods) in Figure 2 show structure breaks with dramatic weights changes, which may imply a guide for analyzing dynamically changing environments. The related analyses are not presented in this paper, available upon request.

Figure 3 presents the distribution forecasting of RV with seven quantiles from the DQMA methods. The RV distributions have heavier and longer right tails when there are larger realized values if the subfigures are compared with the top-left subfigure in Figure 3. Thus, it is reasonable to infer that the stylized feature of heavier right tails also has a clustering effect, similar to that of volatility itself. Surely, distribution forecasting provides much more information than mean forecasting. Another interesting observation is that heavier and longer right tails correspond to the declining period of index if Figure 3 is compared with

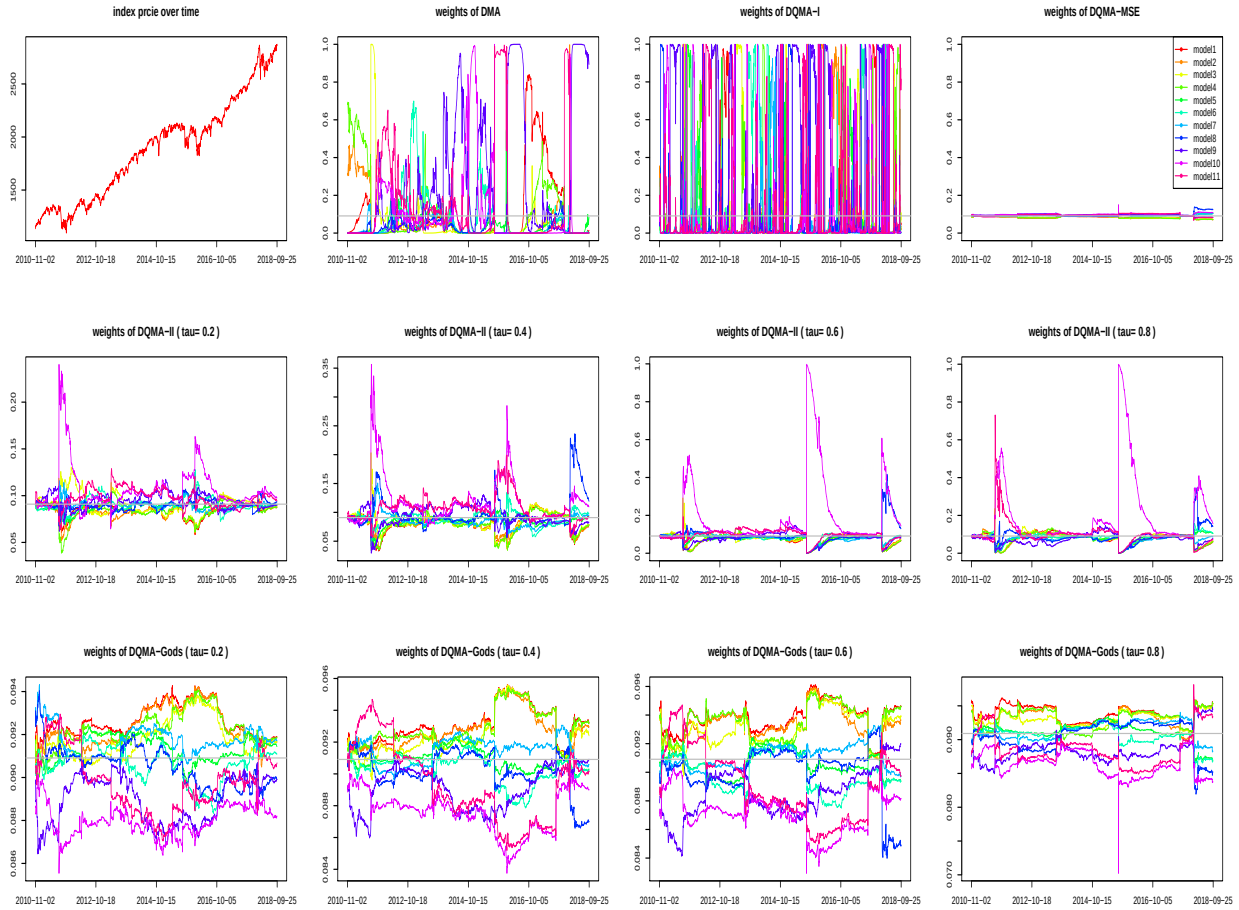

Figure 2: The S&P 500 index and weight evolution of the DMA and DQMA methods. The historical gray line represents the equal averaging weights. The middle and bottom rows are, respectively, the weights of DQMA-II and DQMA-Gods at quantiles τ=(0.2,0.4,0.6,0.8). The colors of the lines in all subfigures are the same as those in the top-right subfigure.

the first subfigure in Figure 2. Thus, the RV distribution is more dispersed during periods of stock market decline.

4 Robustness Checks

First, we further test the quantile predictive power and stability of DQMA methods by comparing them with HARQ models using the MCS procedure (namely, the model confidence set developed by Hansen et al. (2011)) with the asymmetric VaR loss function of González-Rivera et al. (2004), which applies heavier penalties to observations that violate the predicted VaR. As stated by Bernardi et al. (2017), the tests in the above subsection 3.4.1 can fail to

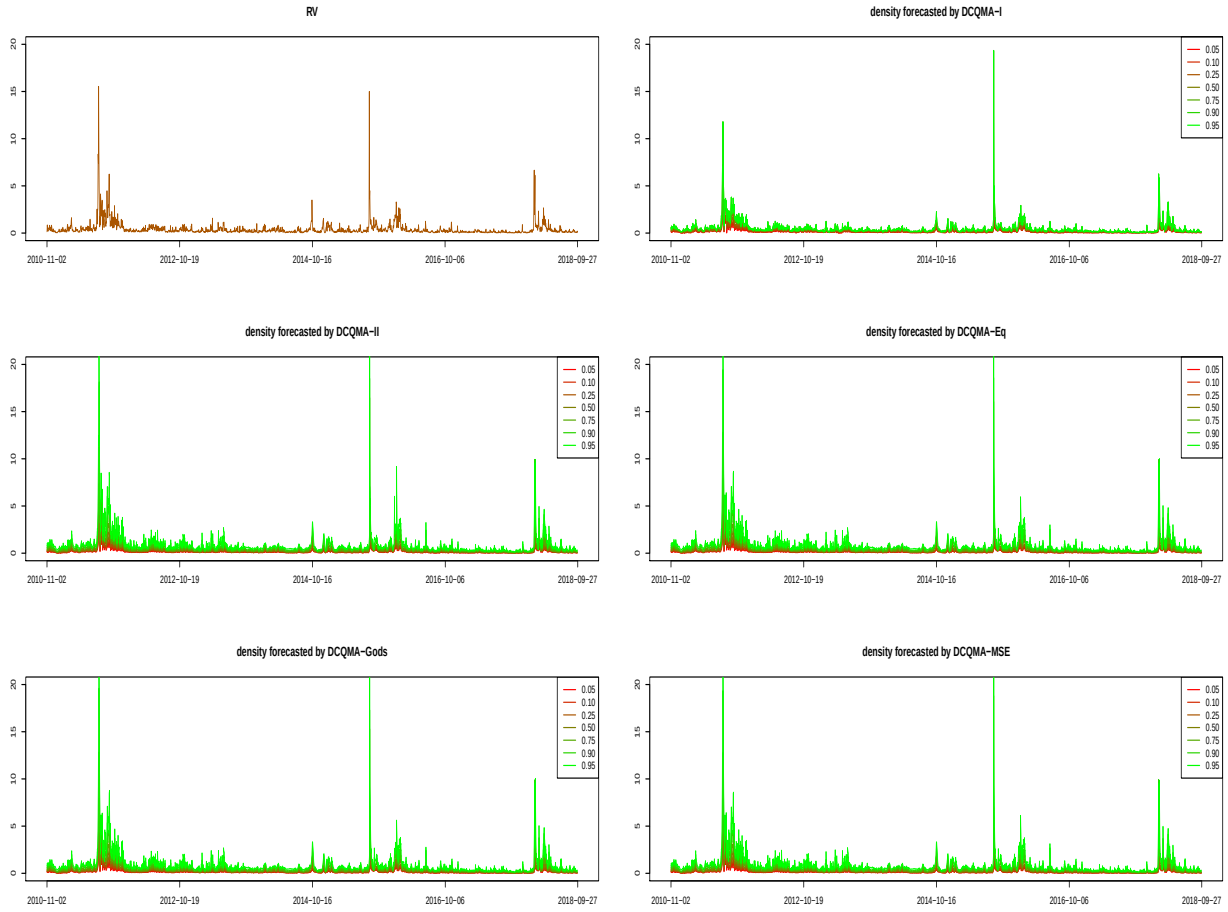

Figure 3: Density forecast by quantile combination method. We draw 7 lines in each subfigure at quantiles $q_\tau, \tau = (0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95)$.

predict accuracy of the delivered quantile estimates and are insensitive to the magnitude of the violation. Moreover, those tests cannot directly compare each of the competing models. Therefore, the MCS procedure is adopted to build the superior set of models (SSM), where the performances of the models are equal at confidence level $(1 - \alpha)$. Actually, Bernardi et al. (2017) also use the MCS procedure to select VaR models and to compare individual VaR models.

Table 8 shows the MCS examination results of nine quantile forecasts from individual HARQ models and DQMA models. Individual quantile regression models can be excluded from the MCS sets, especially for low quantile forecasting: HARQ3 and HARQ4 are eliminated from the SSM in $\tau = 0.25$ and $\tau = 0.5$. Meanwhile, the DQMA methods are always included in the SSM.

Table 8: MCS examination of individual regression models and DQMAs at different quantiles

τ	0.05	0.1	0.25	0.4	0.5	0.75	0.9	0.95
HARQ1	1	1	1	1	0	1	1	1
HARQ2	1	1	1	1	1	1	1	1
HARQ3	1	1	0	1	0	1	1	1
HARQ4	1	1	0	1	0	1	1	1
HARQ5	1	1	1	1	1	1	1	1
HARQ6	1	1	1	1	1	1	1	1
HARQ7	1	1	1	1	1	1	1	1
HARQ8	1	1	1	1	1	1	1	1
HARQ9	1	1	1	1	1	1	1	1
HARQ10	1	1	1	1	1	1	1	1
HARQ11	1	1	1	1	1	1	1	1
DQMA-I	1	1	1	1	1	1	1	1
DQMA-II	1	1	1	1	1	1	1	1
DQMA-Eq	1	1	1	1	1	1	1	1
DQMA-Gods	1	1	1	1	1	1	1	1
DQMA-MSE	1	1	1	1	1	1	1	1

Note: The testing confidence level of MCS is 90%. The value is one if the model is included in SSM and zero otherwise.

Second, we discard the two best or worst models and then examine the performances of the combination strategies. From Table 4 to Table 8, the two best individual models in the whole empirical period are HARQ6 and HARQ11. Therefore, these two models are removed to assess whether the performances of the combining strategies will degrade. Then, the two worst models (HARQ3 and HARQ4) are discarded. The results are shown in Table 9 and Table 10.

Finally, Table 9 shows the goodness of fit of the quantiles after discarding two individual models. The goodness of fit declines slightly when the two best individual models are discarded and increases slightly after discarding HARQ3 and HARQ4. However, all the models are included in the SSM at the 90% confidence level after discarding two models, as shown in Table 10, which means the DQMA method is robust to the included individual models. Therefore, the combination strategies are superior to individual models in terms of the risk of model uncertainty.

5 Conclusions

Using intra-day high-frequency data to study realized volatility has been a common practice in the past two decades. In this paper, we study the RV of the S&P 500, focusing

Table 9: Goodness of fit R_τ^1 of quantile combining strategies for robustness testing

τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
DQMA-I-Dis6-11	0.1523	0.2285	0.2888	0.3317	0.3676	0.3414	0.4023	0.4589	0.5298
DQMA-II-Dis6-11	0.1990	0.2446	0.2914	0.3325	0.3714	0.4020	0.4498	0.4976	0.5601
DQMA-Eq-Dis6-11	0.1986	0.2449	0.2907	0.3305	0.3698	0.3998	0.4449	0.4953	0.5594
DQMA-Gods-Dis6-11	0.1984	0.2449	0.2905	0.3303	0.3697	0.4002	0.4453	0.4957	0.5599
DQMA-MSE-Dis6-11	0.1970	0.2441	0.2909	0.3313	0.3707	0.3960	0.4428	0.4941	0.5586
DQMA-I-Dis3-4	0.1910	0.2474	0.2945	0.3399	0.3748	0.3918	0.4390	0.4848	0.5508
DQMA-II-Dis3-4	0.2027	0.2508	0.2965	0.3370	0.3739	0.4046	0.4523	0.5009	0.5674
DQMA-Eq-Dis3-4	0.2023	0.2509	0.2965	0.3355	0.3737	0.4053	0.4514	0.5001	0.5628
DQMA-Gods-Dis3-4	0.2021	0.2509	0.2965	0.3354	0.3736	0.4058	0.4520	0.5006	0.5634
DQMA-MSE-Dis3-4	0.2033	0.2499	0.2963	0.3358	0.3740	0.4012	0.4489	0.4986	0.5616

Note: Numbers in red indicate that the goodness of fit declines compared with the full combined DQMA with eleven individual models.

Table 10: MCS examinations of DQMA models after discarding two models

τ	0.05	0.1	0.25	0.5	0.75	0.9	0.95
DQMA-I	1	1	1	1	1	1	1
DQMA-II	1	1	1	1	1	1	1
DQMA-Eq	1	1	1	1	1	1	1
DQMA-Gods	1	1	1	1	1	1	1
DQMA-MSE	1	1	1	1	1	1	1
DQMA-I-Dis6-11	1	1	1	1	1	1	1
DQMA-II-Dis6-11	1	1	1	1	1	1	1
DQMA-Eq-Dis6-11	1	1	1	1	1	1	1
DQMA-Gods-Dis6-11	1	1	1	1	1	1	1
DQMA-MSE-Dis6-11	1	1	1	1	1	1	1
DQMA-I-Dis3-4	1	1	1	1	1	1	1
DQMA-II-Dis3-4	1	1	1	1	1	1	1
DQMA-Eq-Dis3-4	1	1	1	1	1	1	1
DQMA-Gods-Dis3-4	1	1	1	1	1	1	1
DQMA-MSE-Dis3-4	1	1	1	1	1	1	1

Note: The testing confidence level of MCS is 90%. The values equal one if the model is included in the SSM and zero otherwise. The suffix Disx-y means discarding the corresponding model x and model y; for example, DQMA-I-Dis6-11 means discarding HARQ6 and HARQ10 from the original eleven models.

on quantile forecasting. Heterogeneity, volatility persistence and leverage effects are well captured by HAR-RV type models, and many extensions of HAR-RV have been reported since the original method proposed in Corsi (2009). However, the conditional quantile or distribution forecasting of RV has received much less attention than mean forecasting. Taking full advantages of nonparametric realized measures and quantile regression, this paper avoids rigid restrictive distributional assumptions on the dynamics of RV, which often poses serious problems to econometric modeling. In fact, the heterogeneous autoregressive quantile model

performs well in capturing the dynamics of the RV conditional distributions; see Žikeš and Baruník (2016). Considering that the forecasting accuracy of each individual HARQ model does not adapt to the full features of a dynamically changing environment and face the risk of model uncertainty, five dynamic quantile model averaging strategies are proposed to address this problem. Furthermore, eleven HARQ models are selected with more than 35 realized measures to study the quantile dynamics and forecasting of RV.

Our empirical examinations indicate clearly that the stylized features of RV are different at different quantiles, with stronger features at high-level quantiles. For example, persistence and leverage effects are much stronger at high-level quantiles than at low-level quantiles. bad volatility and bad jump has stronger impacts on high-level quantiles. These unequal impacts affect the tails of RV and cause heavier right tails. From the forecasting densities of RV obtained by DQMA, we can infer that the heavier right tail also has a clustering effect, similar to volatility itself. However, the effects of jumps are much more complicated and mixed.

More importantly, we evaluate comprehensively the forecasting performances of DQMA methods with several evaluation criteria. To show the advantages of DQMA and HARQ models in modeling the dynamics of RV compared with mean models, we choose two best mean models as benchmarks for comparison. They are the most common popular benchmark competing ARFIMA model and DMA-TVP of Wang et al. (2016), which outperformed eighteen competing combining strategies. As expected, the DQMA and HARQ models outperform DMA-TVP and ARFIMA in both quantile forecasting and interval forecasting. Furthermore, not only do DQMA models demonstrate excellent performances in modeling the quantiles of RV of the S&P 500, but also they are more stable than individual HARQ models and can avoid the risk of model uncertainty. In addition, the dynamic evolution of the combining weights may offer information about the dynamically changing environment, such as structure breaks. In summary, DQMA methods can not only make accurate RV quantile forecasts and capture the dynamics of RV well without rigid distributional assumptions but also avoid the risk of model uncertainty.

References

Aiolfi, M. and Timmermann, A. (2006). Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics*, 135(1-2):31–53.

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

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61(1):43–76.

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

Audrino, F. and Hu, Y. (2016). Volatility forecasting: Downside risk, jumps and leverage effect. *Econometrics*, 4, 8(1):1–24.

Audrino, F. and Knaus, S. D. (2016). Lassoing the har model: A model selection perspective on realized volatility dynamics. *Econometric Reviews*, 35(8-10):1485–1521.

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B*, 64(2):253–280.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2(1):1–37.

Bates, J. M. and Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468.

Berkowitz, J., Christoffersen, P., and Pelletier, D. (2011). Evaluating value-at-risk models with desk-level data. *Management Science*, 57(12):2213–2227.

Bernardi, M., Catania, L., and Petrella, L. (2017). Are news important to predict the value-at-risk? *European Journal of Finance*, 23(6):535–572.

Bollerslev, T., Kretschmer, U., Pigorsch, C., and Tauchen, G. (2009). A discrete-time model for daily s & p500 returns and realized variations: Jumps and leverage effects. *Journal of Econometrics*, 150(2):151–166.

Bollerslev, T., Russell, J., and Watson, M. (2010). *Volatility and Time Series Econometrics: Essays in Honor of Robert Engle*. Oxford University Press.

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference –A Practical Information-Theoretic Approach*. 2nd Edition, Springer, New York.

Burnham, K. P. and Anderson, D. R. (2004). Multimodel inference: understanding aic and bic in model selection. *Sociological Methods & Research*, 33(2):261–304.

Cai, Z. and Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association*, 103(484):1595–1608.

Caporin, M., Rossi, E., and Magistris, P. S. d. (2014). Volatility jumps and their economic determinants. *Journal of Financial Econometrics*, 14(1):29–80.

Chan, F. and Pauwels, L. L. (2018). Some theoretical results on forecast combinations. *International Journal of Forecasting*, 34(1):64–74.

Christensen, K., Oomen, R. C., and Podolskij, M. (2014). Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics*, 114(3):576–599.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841–862.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Corsi, F., Mittnik, S., Pigorsch, C., and Pigorsch, U. (2008). The volatility of realized volatility. *Econometric Reviews*, 27(1-3):46–78.

Corsi, F., Pirino, D., and Renò, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*, 159(2):276–288.

Corsi, F. and Renò, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics*, 30(3):368–380.

Elliott, G. and Timmermann, A. (2004). Optimal forecast combinations under general loss functions and forecast error distributions. *Journal of Econometrics*, 122(1):47–79.

Engle, R. F. and Manganelli, S. (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381.

Fan, J., Yao, Q., and Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, 83(1):189–206.

Giacomini, R. and Komunjer, I. (2005). Evaluation and combination of conditional quantile forecasts. *Journal of Business & Economic Statistics*, 23(4):416–431.

González-Rivera, G., Lee, T.-H., and Mishra, S. (2004). Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of Forecasting*, 20(4):629–645.

Granger, C. W. (1989). Invited review combining forecaststwenty years later. *Journal of Forecasting*, 8(3):167–173.

Granger, C. W. and Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, 3(2):197–204.

Granger, C. W. J., White, H., and Kamstra, M. (1989). Interval forecasting: an analysis based upon arch-quantile estimators. *Journal of Econometrics*, 40(1):87–96.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2):453–497.

Hendry, D. F. and Clements, M. P. (2004). Pooling of forecasts. *Econometrics Journal*, 7(1):1–31.

Hounyo, U., Gonçalves, S., and Meddahi, N. (2017). Bootstrapping pre-averaged realized volatility under market microstructure noise. *Econometric Theory*, 33(4):791–838.

Hsiao, C. and Wan, S. K. (2014). Is there an optimal forecast combination? *Journal of Econometrics*, 178(2):294–309.

Inoue, A., Jin, L., and Rossi, B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics*, 196(1):55–67.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

Koenker, R. and Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94(448):1296–1310.

Koenker, R. and Xiao, Z. (2002). Inference on the quantile regression process. *Econometrica*, 70(4):1583–1612.

Koenker, R. and Xiao, Z. (2004). Unit root quantile autoregression inference. *Journal of the American Statistical Association*, 99(467):775–787.

Koenker, R. and Xiao, Z. (2006). Quantile autoregression. *Journal of the American Statistical Association*, 101(475):980–990.

Koop, G. and Korobilis, D. (2012). Forecasting inflation using dynamic model averaging. *International Economic Review*, 53(3):867–886.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3(2):73–84.

Liu, C. and Maheu, J. M. (2009). Forecasting realized volatility: a bayesian model-averaging approach. *Journal of Applied Econometrics*, 24(5):709–733.

Liu, L. Y., Patton, A. J., and Sheppard, K. (2015). Does anything beat 5-minute rv? a comparison of realized measures across multiple asset classes. *Journal of Econometrics*, 187(1):293–311.

McAleer, M. and Da Veiga, B. (2008). Single-index and portfolio models for forecasting value-at-risk thresholds. *Journal of Forecasting*, 27(3):217–235.

Patton, A. J. and Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics*, 97(3):683–697.

Raftery, A. E., Kárnỳ, M., and Ettler, P. (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics*, 52(1):52–66.

Steel, M. F. J. (2020). Model averaging and its use in economics. *Journal of Economic Literature*, 58(3):644–719.

Stock, J. H. and Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23(6):405–430.

Taylor, J. W. (2005). Generating volatility forecasts from value at risk estimates. *Management Science*, 51(5):712–725.

Taylor, J. W. and Bunn, D. W. (1998). Combining forecast quantiles using quantile regression: Investigating the derived weights, estimator bias and imposing constraints. *Journal of Applied Statistics*, 25(2):193–206.

Timmermann, A. (2006). Forecast combinations. in *Handbook of Economic Forecasting*, Vol. 1:135–196.

Wallis, K. F. (2011). Combining forecasts–forty years later. *Applied Financial Economics*, 21(1-2):33–41.

Wang, Y., Ma, F., Wei, Y., and Wu, C. (2016). Forecasting realized volatility in a changing world: A dynamic model averaging approach. *Journal of Banking & Finance*, 64(1):136–149.

Yang, Y. (2004). Combining forecasting procedures: some theoretical results. *Econometric Theory*, 20(1):176–222.

Yu, K., Lu, Z., and Stander, J. (2003). Quantile regression: applications and current research areas. *Journal of the Royal Statistical Society: Series D*, 52(3):331–350.

Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, 54(4):437–447.

Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli*, 12(6):1019–1043.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472):1394–1411.

Žikeš, F. and Baruník, J. (2016). Semi-parametric conditional quantile models for financial returns and realized volatility. *Journal of Financial Econometrics*, 14(1):185–226.