DOI 10.7603/s40570-016-0002-9

Volume 18, Number 1 – March 2016
China Accounting and Finance Review
中国会计与财务研究
2016 年 3 月 第 18 卷 第 1 期

Realised Volatility Forecasts for Stock Index Futures Using the HAR Models with Bayesian Approaches*

Jiawen Luo[1] and Langnan Chen[2]

Received 5th of March 2015 Accepted 26th of November 2015

Abstract

We investigate the realised volatility (RV) forecasts for the short, mid, and long term by developing the HAR models with Bayesian approaches and employing the high-frequency data of the China Stock Index 300 (CSI300) future for the period from 16 April 2010 to 21 May 2014. We also evaluate the performances of competing models for both in-sample forecasts and out-of-sample forecasts. We find that the proposed HAR-type models with Bayesian approaches capture the time-varying properties of parameters and predictor sets. We also find that the HAR-type models with Bayesian approaches have superior forecast performance for both in-sample forecasts and out-of-sample forecasts as compared with the benchmark HAR-type models.

Keywords: Realised Volatility Forecast, Stock Index Futures, HAR Model, Bayesian Approaches, Time-varying

基于 HAR 模型和贝叶斯方法股指期货已实现波动率的预测

罗嘉雯 陈浪南 中山大学岭南学院

摘要

本文以中国股指期货 2010 年 4 月 16 日到 2014 年 5 月 21 日的高频数据为研究样本，通过构建结合贝叶斯方法的 HAR 模型对中国股指期货的已实现波动率进行短期、中期和长期的预测。同时，本文对不同模型的样本内和样本外预测进行评价。结果表明：结合贝叶斯方法的 HAR 模型能够更好地体现参数和预测变量集的时变特征。此外，与原模型相比，结合贝叶斯方法的 HAR 模型具有更好地样本内和样本外预测效果。

关键词：已实现波动率的预测、股指期货、HAR 模型、贝叶斯方法、时变特征

* We gratefully acknowledge the valuable suggestions made by the Executive Editor of the *China Accounting and Finance Review* and the anonymous referees. This research is supported by the China Social Science Foundation under Grant No. 14ZDA020, the Humanities and Social Sciences Foundation of Chinese Ministry of Education under Grant No. 14YJA7900.
作者衷心感谢《中国财务与会计研究》主编和匿名评审人的建议。本文为国家社科基金重大课题（14ZDA020）、教育部规划基金项目（14YJA7900）、及广东软科学（2013B070206025）资助成果之一。

1 Jiawen Luo, Ph.D candidate, Lingnan College, Sun Yat-sen University, P. R. China; email: 2jialuo@163.com.
2 Langnan Chen, corresponding author, Professor of Finance and Economics, Lingnan College, Sun Yat-sen University, P. R. China; email: lnscln@mail.sysu.edu.cn.

Published online: 21 March 2016

I. Introduction

Accurate forecast of volatility is central for asset pricing, portfolio selection, and risk management. The realised volatility (RV) proposed by Andersen and Bollerslev (1998) is one of the most popular *ex post* measures of return volatility and is widely used in volatility forecasts. In contrast to the traditional volatility measures such as square return, conditional variance with (generalised) autoregressive conditional heteroskedasticity ((G)ARCH-type) models, and stochastic volatility (SV), RV is more efficient as it is constructed with high-frequency data that are more informative. In addition to RV, a number of realised measures of high-frequency data have been developed in the recent literature, such as the realised bi-power variation (BPV) of Barndorff-Neilsen and Shephard (2004), the threshold bi-power variation (TBPV) of Corsi *et al.* (2010), and the semi-variation of Barndorff-Nielsen *et al.* (2010).

For RV forecasts, Koopman *et al.* (2005) construct the SV-RV model and the autoregressive fractionally integrated moving average (ARFIMA)-RV model by incorporating RV into the conventional SV models and the ARFIMA models, respectively. Engle and Gallo (2006) present the multiplicative error model (MEM) that includes the RV in the GARCH equation. The heterogeneous autoregressive regression (HAR) models proposed by Corsi (2009) are a recent development in modelling RV. The HAR model for RV forecast is constructed by incorporating the past daily, weekly, and monthly volatility components into the AR model based on the heterogeneous market hypothesis and the HARCH model of Müller *et al.* (1997). The HAR model's tractable estimation, flexible structure, and superior forecast performance have resulted in several extended models. Andersen *et al.* (2007) find that the discontinuous part in RV contributes little to the volatility forecast. They separate RV into a continuous component (C) and a discontinuous component (jumps, J) and construct the HAR-CJ model by incorporating them into a conventional HAR model. As an extension, Corsi *et al.* (2010) suggest a HAR-TCJ model by introducing a threshold in estimating the variations and jumps, which provides a more accurate forecast of the discrete part in RV. More recently, Patton and Sheppard (2015) construct the HAR-ΔJ model by incorporating a signed jump variation in the HAR model, where the RV is decomposed into a good volatility and a bad volatility so that the leverage effect in volatility forecasts can be taken into account. Moreover, Corsi and Renò (2012) propose the LHAR model by considering the leverage effect with a long-range dependence in the HAR model. Bandi and Renò (2012) provide a nonparametric estimation in the continuous-time stochastic volatility model with both jumps in returns and variance, making it feasible to identify the time-varying leverage effects in the HAR model.

However, most of the forecast models above are based on constant parameters and invariable predictor sets. Due to exogenous shocks, such as policy changes and financial crises, volatility usually exhibits different patterns during different periods. The parameters in the RV forecast models might change over time. In addition, there might be a model risk associated with specifying a single model with invariable predictors over time as there is considerable model uncertainty. In other words, the ranking of predictability for a predictor can vary over the forecast horizons. Thus, forecasting volatility without taking into account the time-varying properties of parameters and predictors will result in a bias.

More recently, Bayesian approaches have been employed to provide a more flexible way of volatility forecast. Wright (2008) proposes a Bayesian model averaging (BMA) approach to account for past and future structural breaks. Liu and Maheu (2009) utilise the BMA approach to forecast RV and suggest that compared with the benchmark models, the BMA approach improves both point forecasts and density forecasts. Raftery *et al.* (2010) improve the BMA approach by introducing forgetting parameters into the estimation and also propose new types of BMA approaches: dynamic model averaging (DMA) and dynamic model selection (DMS). Koop and Korobilis (2012) use DMA and DMS to forecast inflation for both one-step and multi-step ahead forecasts.

To capture the time-varying properties of parameters and predictor sets, we develop HAR-type models with the Bayesian approaches and use the proposed models to forecast the RV of stock index futures for one-step and multi-step ahead forecasts. We also evaluate the in-sample forecast performances on the basis of the Mincer-Zarnowitz regression (MZ-R^2), the mean square error (MSE) loss function, the quasi-likelihood (QLIKE) loss function, and the sum of log predicted likelihood and the out-of-sample forecast performances on the basis of the loss function of Patton (2011) and the model confidence set (MCS) of Hansen *et al.* (2011) among the competing models.

The remainder of this paper is organised as follows: section II develops the HAR-type models with Bayesian approaches; section III describes the high-frequency data and statistics; section IV presents the in-sample forecast and out-of-sample forecast results; and section V concludes the paper.

II. Methodologies

2.1 Realised Volatility

Realised volatility is defined as

$$RV_{t+1}(\delta)=\sum_{j=1}^{1/\delta} r_{t+j\delta,\delta}^{2}\ , \tag{1}$$

where δ is the sample frequency of the RV and $r_{t+j\delta,\delta}$ is the 5-minute frequency returns calculated by $r_{t+j\delta,\delta}=100\times(\log P_{t+j\delta}-\log P_{t+(j-1)\delta})$.

Barndorff-Neilsen and Shephard (2004) present a jump-robust measure of RV called the realised BPV to obtain a robust estimate of jumps. BPV is calculated as

$$BPV_{t+1}(\delta) = \mu_1^{-2}\sum_{j=2}^{1/\delta}\left|r_{t+j\delta,\delta}\right|\left|r_{t+(j-1)\delta,\delta}\right|, \tag{2}$$

where $\mu_1 = \sqrt{1/\pi} = E(|\text{Z}|)$ is the mean of the absolute value of random value Z, Z~*N*(0,1). When $\delta \to 0$, $BPV_{t+1}(\delta) \to \int_t^{t+1}\sigma_s^2 ds$, where σ_s^2 is a diffusive càdlàg process based on the data generation process of price as detailed in Appendix A. Corsi *et al.* (2010) propose an alternative estimate of the integrated powers of volatility called the threshold bi-power variation (TBPV) which provides less biased estimates regarding the standard multi-power variation of the continuous quadratic variation in finite samples. TBPV is calculated as

$$TBPV_t(\delta) = \mu_1^{-2}\sum_{j=2}^{1/\delta}\left|r_{t+(j-1)\delta,\delta}\right|\left|r_{t+j\delta,\delta}\right|\text{I}\left\{r_{t+(j-1)\delta,\delta}^2 \le \vartheta_{(j-1)\delta} \cup r_{t+j\delta,\delta}^2 \le \vartheta_{j\delta}\right\} \tag{3}$$

where I{ } is the indication function, $\vartheta_{j\delta} = c_\vartheta^2 \hat{V}_j$, c_ϑ is the threshold-adjust constant, and $\hat{V}_j$ is the non-parameter recursive filter for calculating the partial variance. On the basis of Corsi *et al.* (2010), we set $c_\vartheta = 3$. We calculate the C_Z_t and C_TZ_t statistics of Barndorff-Neilsen and Shephard (2006) and Corsi *et al.* (2010), respectively.

$$C\_Z_t = \frac{(RV_t - BPV_t)/RV_t}{\sqrt{((\frac{\pi}{2})^2 + \pi - 5)\delta\max(1, \frac{TriPV_t}{BPV_t^2})}}$$

$$C\_TZ_t = \frac{(RV_t - TBPV_t)/RV_t}{\sqrt{((\frac{\pi}{2})^2 + \pi - 5)\delta\max(1, \frac{TTriPV_t}{TBPV_t^2})}},$$

where $TriPV_t^\delta = \delta^{-1}\mu_{4/3}^{-3}\sum_{j=3}^{1/\delta}\prod_{k=1}^{3}\left|r_{t+(j-1+k)\delta,\delta}\right|^{\frac{4}{3}}$ and

$TTriPV_t^\delta = \delta^{-1}\mu_{4/3}^{-3}\sum_{j=3}^{1/\delta}\prod_{k=1}^{3}\left|r_{t+(j-1+k)\delta,\delta}\right|^{\frac{4}{3}}\text{I}\left\{r_{t+(j-1+k)\delta,\delta}^2 \le \vartheta_{(j-1+k)\delta}\right\}$.

Both C_Z_t and C_TZ_t follow a standard normal distribution. The RV is divided into two components: a continuous component and a discontinuous component (jumps).

$$\hat{J}_t = I(C\_Z_t > \Phi_\alpha)\times(RV_t - BPV_t)^+, \qquad \hat{C}_t = RV_t - \hat{J}_t$$
$$\hat{TJ}_t = I(C\_TZ_t > \Phi_\alpha)\times(RV_t - TBPV_t)^+, \qquad T\hat{C}_t = RV_t - T\hat{J}_t \tag{4}$$

Barndorff-Nielsen *et al.* (2010) use the semi-variation to separate the RV into a positive part and a negative part by considering a leverage effect:

$$RS_t^- = \sum_{j=1}^{1/\delta} r_{t+j\delta,\delta}^2 I\left[r_{t+j\delta,\delta} < 0\right], \qquad RS_t^+ = \sum_{j=1}^{1/\delta} r_{t+j\delta,\delta}^2 I\left[r_{t+j\delta,\delta} > 0\right], \tag{5}$$

where $RV = RS^- + RS^+$. Patton and Sheppard (2015) define the signed jump variation as:

$$\Delta J_t = RS_t^+ - RS_t^-.$$

2.2 HAR-type Models

The realised measures are defined as $\overline{RM_{t,h}} = \frac{1}{h}\sum_{j=1}^{h} RM_{t-j}$, where RM represents realised measures such as *RV, C, J, TC, TJ,* and *BPV* in (6)-(9). $\overline{RM_{t,1}}$, $\overline{RM_{t,5}}$, and $\overline{RM_{t,22}}$ are the average estimates of the past 1, 5, and 22 trading days, corresponding to the daily, weekly, and monthly measures of RV, respectively. The proposed forecast models are based on four types of HAR models, namely the traditional HAR models (Corsi, 2009), the HAR-CJ model (Andersen *et al.*, 2007), the HAR-TCJ model (Corsi *et al.*, 2010), and the HAR-ΔJ model (Patton and Sheppard, 2015), which are defined as follows:

HAR: $RV_t = a_0 + a_d \overline{RV}_{t,1} + a_w \overline{RV}_{t,5} + a_m \overline{RV}_{t,22} + u_t$ (6)

HAR-CJ: $RV_t = a_0 + a_d \overline{C}_{t,1} + a_w \overline{C}_{t,5} + a_m \overline{C}_{t,22} + a_{dJ} \overline{J}_{t,1} + a_{wJ} \overline{J}_{t,5} + a_{mJ} \overline{J}_{t,22} + u_t$ (7)

HAR-TCJ: $RV_t = a_0 + a_d \overline{TC}_{t,1} + a_w \overline{TC}_{t,5} + a_m \overline{TC}_{t,22} + a_{dJ} \overline{TJ}_{t,1} + a_{wJ} \overline{TJ}_{t,5} + a_{mJ} \overline{TJ}_{t,22} + u_t$ (8)

HAR-ΔJ: $RV_t = a_0 + a_{\Delta J} \overline{\Delta J}_{t,1} + a_d \overline{BPV}_{t,1} + a_w \overline{RV}_{t,5} + a_m \overline{RV}_{t,22} + u_t$ (9)

2.3 HAR-type Models with Bayesian Approaches

Assume that X_t is a predictor set with *m* predictors and that we have *K* sub-models which are characterised by having different combinations of predictors, where $K = 2^m$.[3] We denote the predictors in each sub-model as $X_t^{(k)}$, where $k = 1,2,\cdots,K$. Consider a state space forecast model with the time-varying coefficients:

$$\begin{aligned} RV_t &= X_t^{(k)} \beta_t^{(k)} + \varepsilon_t^{(k)} \\ \beta_t^{(k)} &= \beta_{t-1}^{(k)} + \eta_t^{(k)} \end{aligned} \quad (10)$$

where both $\varepsilon_t^{(k)}$ and $\eta_t^{(k)}$ are normally distributed: $\varepsilon_t^{(k)} \sim N(0, H_t^{(k)})$ and $\eta_t^{(k)} \sim N(0, Q_t^{(k)})$. Denote that $L_t \in \{1,2,\cdots,K\}$ is the model applied in period *t*, $\mathrm{B}_t = \{\beta_t^{(1)}{}',\cdots,\beta_t^{(K)}{}'\}$ is a set of coefficients, and $RV^t = \{RV_1,\cdots,RV_t\}'$ is the information set at time *t*.

When *m* is large, the number of sub-models *K* is large and it is burdensome to employ the full Bayesian approach for computation. Thus, the application of a Kalman filter helps to reduce the computation workload. Based on Raftery *et al.* (2010), two parameters, *λ* and *α*, which are both forgetting factors, are involved in the estimation models. *λ* and *α* are set slightly less than 1 such that $\lambda, \alpha \in [0.95, 1]$. The role of forgetting factors is to reduce the weight of information for a more distant period as we make observations at time *j* in the past, suggesting a gradual evolution of coefficients. For example，if *λ* = 0.99, the observations at time 20 ahead receive approximately 80% as much weight as those for the last period, while if *λ* = 0.95, the observations at time 20 ahead receive nearly 35% as much weight as those for the last period. In addition, the only distinction between the DMA-HAR-type model and

[3] For the special case where no predictor is incorporated in the forecast model, $K = 2^m$.

the BMA-HAR-type model is that we set $\lambda = \alpha = 1$ for the latter, suggesting that the past information is assigned equal weight and the effect of forgetting factors is eliminated.

For a one-model case, the updating process for coefficients is shown in (B3)-(B4) in Appendix B. For a multi-model case, we denote $L_t = k$ if the *k-th* sub-model M_k is involved in time *t*. The updating process in the multi-model case is similar to that in the one-model case. By denoting $\Sigma_{t|t-1} = \Sigma_{t-1|t-1} + Q_t$, the evolution of coefficients is as follows:

$$\beta_{t-1} \mid L_{t-1} = k, RV^{t-1} \sim N(\hat{\beta}_{t-1}^{(k)}, \Sigma_{t-1|t-1}^{(k)}) \tag{11}$$

$$\beta_t \mid L_t = k, RV^{t-1} \sim N(\hat{\beta}_{t-1}^{(k)}, \Sigma_{t|t-1}^{(k)}) \tag{12}$$

$$\beta_t \mid L_t = k, RV^t \sim N(\hat{\beta}_t^{(k)}, \Sigma_{t|t}^{(k)}) \tag{13}$$

(11) and (13) are respectively the conditional distributions of $\beta_{t-1}^{(k)}$ and $\beta_t^{(k)}$, which are approximated by normal distributions, and (12) is the parameter prediction equation. The updating of $\hat{\beta}_t$ and $\Sigma_{t-1|t-1}$ in (11)-(13) is detailed in (B5) and (B6) in Appendix B.

By incorporating the forgetting parameter λ, we have $Q_t = (\lambda^{-1} - 1)\Sigma_{t-1|t-1}$, $0 < \lambda < 1$, where

$$\Sigma_{t|t-1} = \frac{1}{\lambda}\Sigma_{t-1|t-1}. \tag{14}$$

The recursive forecast of RV_t in the sub-model M_k can be done on the basis of the predictive distribution via Kalman filtering:

$$RV_t \mid RV^{t-1} \sim N(X_t^{(k)}\hat{\beta}_{t-1}^{(k)}, \mathrm{H}_t^{(k)} + X_t^{(k)}\Sigma_{t|t-1}^{(k)}X_t^{(k)\prime}), \tag{15}$$

where $\mathrm{H}_t^{(k)}$ is an important middle parameter in the DMA approach. The estimation of the DMA approach is detailed in Appendix B.

Raftery *et al.* (2010) introduce another forgetting factor *α*, which has the same function as *λ*, in the updating function of the inclusion probability of each sub-model $\pi_{t|t-1,k}$. The updating procedure for the inclusion probabilities of the sub-models is as follows:

$$\pi_{t|t-1,k} = \frac{\pi_{t-1|t-1,k}^{\alpha}}{\sum_{l=1}^{K}\pi_{t-1|t-1,l}^{\alpha}} \tag{16}$$

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k}\, p_k(RV_t \mid RV^{t-1})}{\sum_{l=1}^{K}\pi_{t|t-1,l}\, p_l(RV_t \mid RV^{t-1})}, \tag{17}$$

where $p_l(RV_t \mid RV^{t-1})$ is the prediction density for the *l-th* model conditional on the previous information defined in (15). The derivation of $\pi_{t|t-1,k}$ is detailed in Appendix B. As the inclusion probabilities $\pi_{t|t-1,k}$ of the sub-models evolve over time, the predictors can be considered as the variables with their time-varying weighted average inclusion probabilities. Thus, we can assign more weight to the predictors with higher inclusion probabilities.

Within the BMA and DMA, we obtain the forecast values of RV by calculating the weighted average of all the sub-models' forecast values, while the DMS proceeds by selecting the single model with the highest value of inclusion probabilities at each point in time and simply using it for forecast. The d-step ahead forecast of RV_t is computed by the weighted average of the estimated $R\hat{V}_t^{(k)}$ in each sub-model with the inclusion probability $\pi_{t-d+1|t-d,k}$.

$$RV_t^{DMA/BMA} = \sum_{k=1}^{K} \pi_{t-d+1|t-d,k} R\hat{V}_t^{(k)} = \sum_{k=1}^{K} \pi_{t-d+1|t-d,k} X_{t-d+1}^{(k)} \hat{\beta}_{t-d}^{(k)}$$

$$RV_t^{DMS} = X_{t-d+1}^{(\hat{k})} \hat{\beta}_{t-d}^{(\hat{k})} \text{ , where, } \left\{ \hat{k} : \pi_{t-d+1|t-d,k} = \max\left\{ \pi_{t-d+1|t-d,1}, \cdots, \pi_{t-d+1|t-d,l}, l = 1, \cdots, \mathrm{K} \right\} \right\} \quad (18)$$

In this paper, we construct 12 combination models by incorporating the DMA, DMS, and BMA approaches into the four HAR-type models respectively. Regarding the four HAR-type models (HAR model, HAR-CJ model, HAR-TCJ model, and HAR-ΔJ model), the predictor sets $X_t^{(k)}$ in (10) are $\left\{ \overline{RV}_{t,1}, \overline{RV}_{t,5}, \overline{RV}_{t,22} \right\}$, $\left\{ \overline{C}_{t,1}, \overline{C}_{t,5}, \overline{C}_{t,22}, \overline{J}_{t,1}, \overline{J}_{t,5}, \overline{J}_{t,22} \right\}$, $\left\{ \overline{TC}_{t,1}, \overline{TC}_{t,5}, \overline{TC}_{t,22}, \overline{TJ}_{t,1}, \overline{TJ}_{t,5}, \overline{TJ}_{t,22} \right\}$, and $\left\{ \overline{\Delta J}_{t,1}, \overline{BPV}_{t,1}, \overline{RV}_{t,5}, \overline{RV}_{t,22} \right\}$, respectively. Considering the long memory property in the serials, we specify weak forgetting factors $\lambda = \alpha = 0.99$.

2.4 Performance Evaluation

We use the MZ-R^2, the MSE loss function, the QLIKE loss function, and the sum of log predicted likelihood to evaluate the in-sample forecast performances. The MSE loss function and the QLIKE loss function are two special cases of loss functions in Patton (2011) which are robust to the presence of noise in the volatility proxy.

We also use the loss functions and the MCS to evaluate the out-of-sample forecast performances. Patton (2011) proposes a class of loss functions that are robust to the noise of the volatility proxy and are homogeneous of degree *b*+2:

$$L(RV_{True,t}, RV_{Forecast,t}, b) = \begin{cases} \frac{1}{\mathrm{T}_2} \sum_{t=\mathrm{T}_1+1}^{T} \left[\frac{1}{(b+1)(b+2)} (RV_{True,t})^{b+2} - (RV_{Forecast,t})^{b+2}) - \frac{1}{(b+1)} (RV_{Forecast,t})^{b+1} (RV_{True,t} - RV_{Forecast,t}) \right], b \neq -1, -2 \\ \frac{1}{\mathrm{T}_2} \sum_{t=\mathrm{T}_1+1}^{T} (RV_{Forecast,t} - RV_{True,t} + RV_{True,t} \ln \frac{RV_{True,t}}{RV_{Forecast,t}}), b = -1 \\ \frac{1}{\mathrm{T}_2} \sum_{t=\mathrm{T}_1+1}^{T} (\frac{RV_{True,t}}{RV_{Forecast,t}} - \ln \frac{RV_{True,t}}{RV_{Forecast,t}} - 1), b = -2 \end{cases} \quad (19)$$

On the basis of Patton (2011), when $b = 0$ and $b = -2$, the loss functions degenerate to two conventional loss functions: the MSE and the QLIKE, respectively. In this paper, we also use the homogeneous robust loss function when b = -1 and the positive robust loss function when b = 1 to evaluate and compare the various forecast models, where the homogeneous robust loss penalises the under-prediction of volatility more heavily while the positive robust loss penalises the over-prediction of volatility more heavily.

In addition, we employ the MCS developed by Hansen *et al.* (2011) to test the significance of forecast performances among various competing models. The MCS procedure aims at selecting a set of models with the best forecast performances from a set of candidate forecast models $\mathcal{M}_0 = \left\{\mathcal{M}^i, i = 1, \cdots, \mathrm{M}\right\}$. By defining the relative performance variable $d_{ij,t} = L_{i,t} - L_{j,t}$, for all $i, j \in M_0$ and $\mu_{ij} = E(d_{ij,t})$, where L is a specified loss function $L_{i,t} = L(RV_t^{true}, RV_t^{forecast})$, model $\mathcal{M}^i$ is preferred to model $\mathcal{M}^j$ if $\mu_{ij} < 0$. The set of superior models is defined as $M^* \equiv \{i \in M_0 : \mu_{ij} \le 0$ for all $j \in M_0\}$. The null hypothesis in the MCS approach is $H_{0,\mathcal{M}} : \mu_{ij} = 0 \;\; for\ all\ i, j \in \mathcal{M}$, where $\mathcal{M} \in \mathcal{M}_0$. Initially, let $\hat{\mathcal{M}}_{1-\alpha} = \mathcal{M}_0$. The MCS algorithm is constructed with an equivalence test and an eliminating rule as follows:

Step 1: Test $H_{0,\mathcal{M}}$ on the basis of an equivalence test at the significance level of *α*. Hansen *et al.* (2011) present three different tests of statistics for the equal predictive accuracy (EPA) hypothesis. We employ two of them: the range statistics T_R and the semi-quadratic statistics T_{SQ}. Both types of EPA test statistics are based on the following *t*-statistics:

$$t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\hat{\mathrm{var}}(\bar{d}_{ij})}} \quad \text{for } \; i, j \in \mathrm{M}, \tag{20}$$

where $\bar{d}_{ij} = \frac{1}{N}\sum_{t=1}^{N} d_{ij,t}$, N is the length of the forecast period, and $\hat{\mathrm{var}}(\bar{d}_{ij})$ is an estimate of $\mathrm{var}(\bar{d}_{ij})$ obtained by using the stationary block bootstrap of Politis and Romano (1994). The *t*-statistics t_{ij} provide scaled information on the average difference in the forecast quality of model *i* and model *j.* The range statistics T_R and the semi-quadratic statistics T_{SQ} are given by

$$T_R = \max_{i,j \in \mathrm{M}} |t_{ij}| = \max_{i,j \in \mathrm{M}} \frac{\left|\bar{d}_{ij}\right|}{\sqrt{\hat{\mathrm{var}}(\bar{d}_{ij})}}$$

$$T_{SQ} = \sum_{i,j \in \mathrm{M}} t_{ij}^2 = \sum_{i,j \in \mathrm{M}} \frac{(\bar{d}_{ij})^2}{\hat{\mathrm{var}}(\bar{d}_{ij})} \tag{21}$$

Step 2: If $H_{0,\mathcal{M}}$ is accepted, set $\hat{\mathcal{M}}^*_{1-\alpha} = \hat{\mathcal{M}}_{1-\alpha}$; otherwise, use the elimination rule to remove a model from $\hat{\mathcal{M}}_{1-\alpha}$ and go back to Step 1.

In general, given the significance level *α* is fixed at each step, $\hat{\mathcal{M}}^*_{1-\alpha}$ contains the best forecast models from $\mathcal{M}_0$ with (1−*α*) confidence.

III. Data

We employ the 5-minute data of the China Stock Index 300 (CSI 300) future from 16 April 2010 (the first day when the CSI 300 future was listed on the market) to 21 May 2014, giving a total of 991 trading days; the total number of observations is 53,514. The data are from the Wind Database. The intraday trading period of the CSI 300 future is from 9:15 until 15:15, and the total number of intraday observations is 54. We first calculate the intraday 5-minute log return *r*. The realised volatility *RV* is obtained by accumulating the intraday returns. In the recent literature, some scholars propose the overnight RV estimator (Ahoniemi and Lanne, 2013; Koopman *et al*., 2005; Martens, 2002) so that the RV can be scaled with the off-trading hours information. There are three reasons why we only adopt the trading hour data to compute the realised variances. First, although the incorporation of overnight return yields a more complete analysis, it will increase the model complexity. In fact, the exclusion of overnight returns is the common practice in the HAR-type model (Andersen *et al.*, 2007; Corsi *et al.*, 2010; Bandi and Reno, 2012; Xu and Perron, 2014; etc.). Second, Hansen and Lunde (2006) demonstrate that the overnight return is far more volatile than the intraday 5-minute returns and will bring extra noise. Tsiakas (2008) suggests that trading hour returns have a different data generating process from overnight returns. Third, as the overnight volatility is proportional to the daily volatility, it might not have an effect on the overall results.

We then calculate the realised *BPV* of Barndorff-Neilsen and Shephard (2004) and the *TBPV* of Corsi *et al.* (2010). We further obtain the continuous components and jumps by employing two types of statistics, C_Z_t and C_TZ_t, which are detailed in Part 2. Finally, we calculate the realised semi-variation RS^+ and RS^- and derive the signed jump component ΔJ on the basis of Barndorff-Nielsen *et al.* (2010) and Patton and Sheppard (2015).

Table 1 presents the descriptive statistics of the data, including the intraday average return *r*, the realised volatility *RV*, the realised bi-power variation *BPV,* the threshold bi-power variation *TBPV,* and the realised semi-variations RS^+ and RS^-. As shown in Table 1, all the realised estimators exhibit the features of high peak and fat tail. The JB statistics of all the variables significantly reject the null hypothesis of being normally distributed. The results of the Ljung-Box statistics indicate that all the variables exhibit significant autocorrelation and all the realised estimators have the property of long memory. Moreover, the significance levels of the ADF statistics indicate that all the variables belong to a stationary time series.

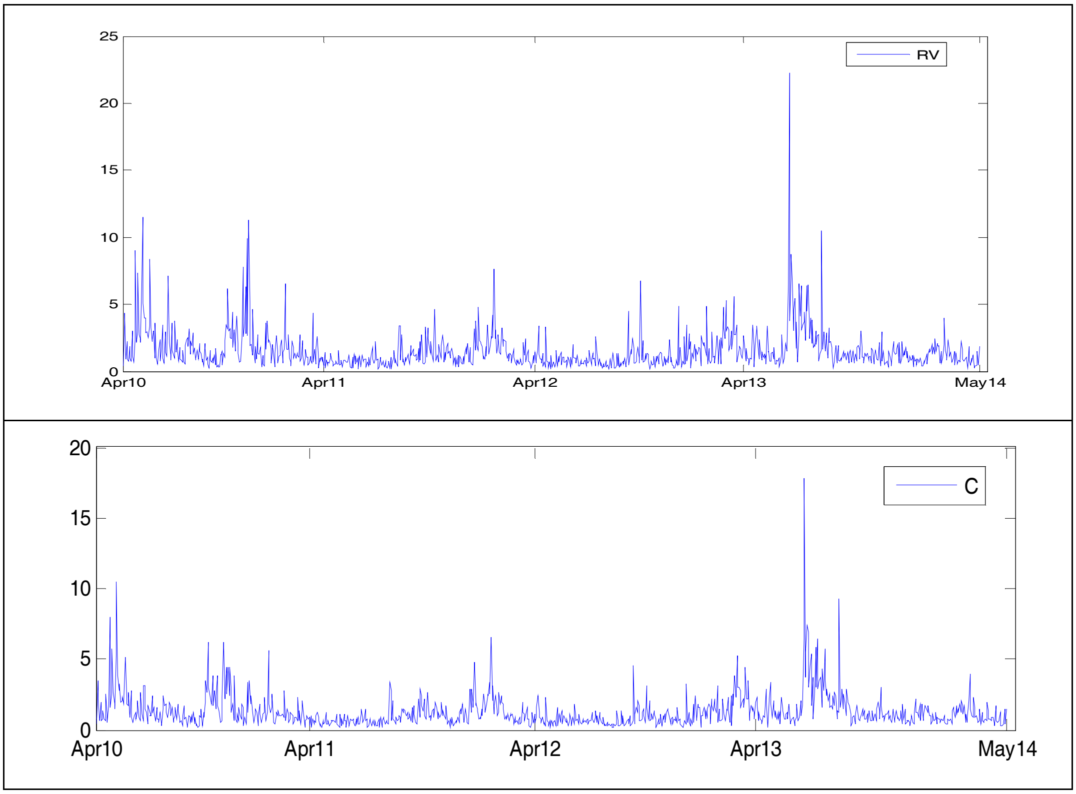
Figure 1 shows the patterns of the calculated continuous components (*C* and *TC*) and the jump components (*J* and *TJ*) of the RV by employing two types of statistics: C_Z_t and C_TZ_t. The first three graphs in Figure 1 show that the continuous components (*C*) departed from the RV exhibit a smoother pattern than the RV, and the threshold continuous component (*TC*) has an even smoother pattern than the continuous components (*C*). The last three graphs in Figure 1 indicate that all the jump components, including the jump with or without the threshold effect (*J* and *TJ*) as well as the signed jump components, exhibit the features of volatility clustering and jumps in particular periods.

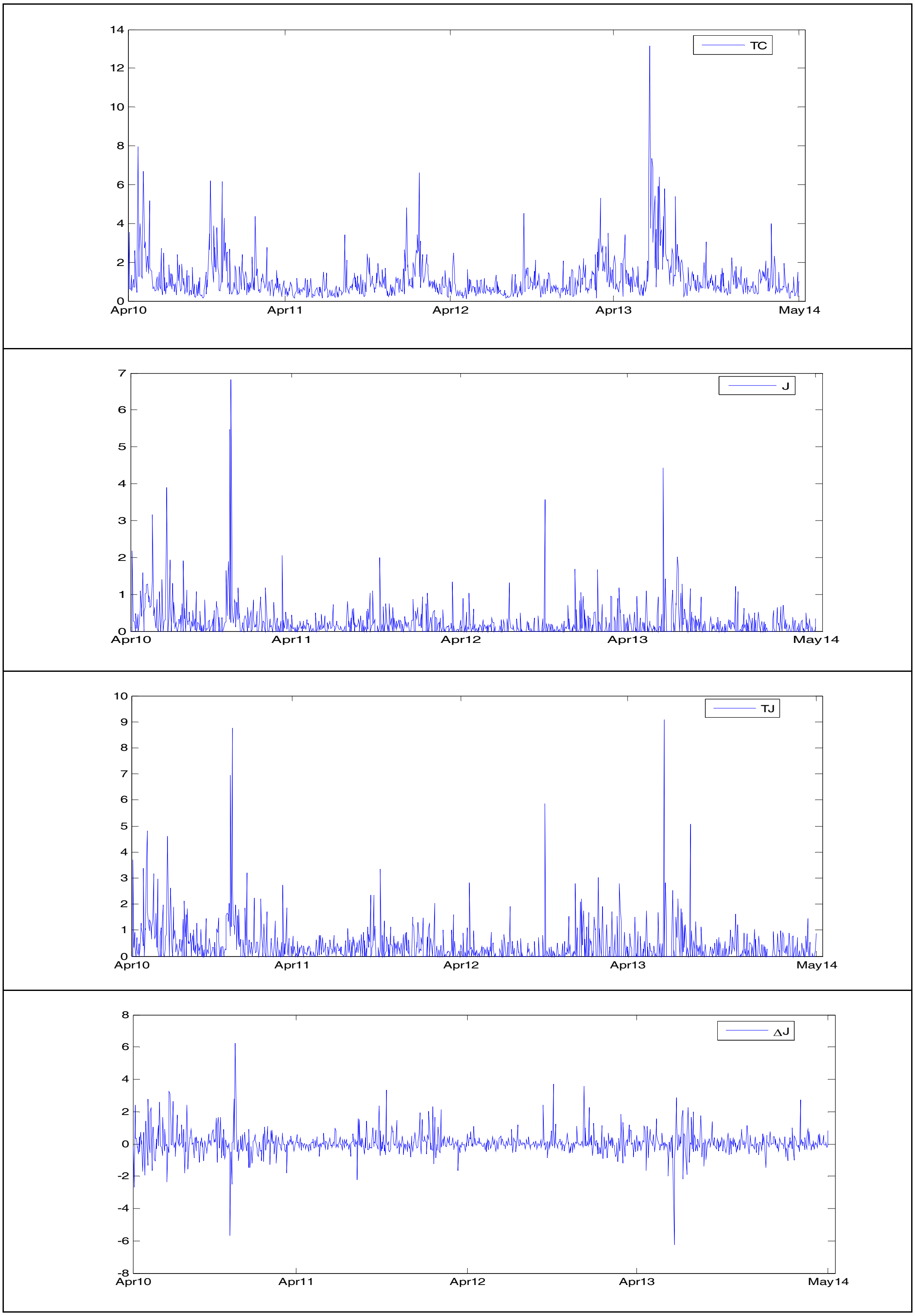
Table 1 Summary Statistics

Variable	R	RV	BPV	TBPV	RS^-	RS^+
Mean	-9.13e-04	1.55	1.30	1.05	0.75	0.80
Standard Deviation	0.18	1.52	1.22	1.02	0.80	0.90
Skewness	0.17	4.69	4.63	4.09	6.57	3.54
Kurtosis	18.07	44.84	44.20	31.83	88.61	27.95
JB statistic	506700***	75400***	73200***	36900***	307850***	15285***
Ljung-Box,Q(5)	43.07***	650.91***	757.73***	786.80***	473.95***	374.49***
Ljung-Box,Q(10)	50.50**	926.87***	1077.60***	1144.11***	692.01***	520.21***
Ljung-Box,Q(20)	63.72	1234.04***	1432.51***	1542.70***	956.52***	670.14***
ADF	-166.10***	-13.10***	-12.88***	-12.48***	-14.36***	-14.46***

Note: Table 1 presents the summary statistics of the variables. *r* is the intraday average return, RV is the realised volatility, BPV is the realised bi-power variation, TBPV is the threshold bi-power variation, and RS^+ (RS^-) is the positive (negative) realised semi-variation. *** denotes a significance level of 1%, ** denotes a significance level of 5%, and * denotes a significance level of 10%.

Figure 1 Continuous Components and Jump Components of the Realised Volatility





Note: Figure 1 shows the time-varying characteristics of the major variables. RV is the realised volatility, C is the continuous component, TC is the threshold continuous component, J is the jump component, TJ is the threshold jump component, and dJ is the sign-jump component.

IV. Realised Volatility Forecasts and Evaluation

4.1 In-Sample Forecasts Using the HAR-type Models

We conduct the in-sample forecasts for different forecast steps by employing the four HAR models mentioned above and compare the forecast performances in terms of three types of statistics: the MZ-R^2,[4] the MSE, and the QLIKE loss function. The coefficients of the HAR models are estimated by ordinary least square, and the t-statistics are calculated with Newey-West HAC (Heteroskedasticity and Autocorrelation Consistent). As shown in Table 2, almost all the coefficients of the HAR-ΔJ and the HAR models are significant at the 10% confidence interval both for one-step and multi-step ahead forecasts, indicating the good in-sample forecast performances of these two models. The HAR-CJ model has significant coefficients for all the continuous components for one-step and five-step ahead forecasts, while all the jump components are insignificant for all forecast steps. Except for the lagged weekly volatility for the one-step ahead forecast and the lagged daily and weekly volatilities for the five-step ahead forecast, all the other coefficients in the HAR-TCJ model are insignificant.

From the results of the three types of statistics, the MZ-R^2, the MSE, and the QLIKE, the HAR-ΔJ model performs best while the HAR model performs worst for all forecast steps for the in-sample forecast. The HAR-TCJ model ranks second for the one-step and 22-step ahead forecasts, while the HAR-CJ ranks second for the five-step ahead forecast.

Table 2 In-Sample Forecast Results Using the HAR-type Models

	h = 1							
Models	HAR-ΔJ		HAR		HAR-CJ		HAR-TCJ	
	coefficient	t-value	coefficient	t-value	coefficient	t–value	coefficient	t-value
a_0	0.2766^{***}	3.6758	0.3069^{***}	3.8404	0.2661^{***}	3.6756	0.2846^{***}	4.0386
a_d	0.1952^{**}	2.4996	0.1174^{***}	2.0522	0.1845^{*}	1.5816	0.1656	1.1258
a_w	0.4675^{***}	5.4665	0.4792^{***}	5.4276	0.5173^{***}	5.1339	0.5782^{***}	4.9989
a_m	0.1890^{**}	2.1737	0.1957^{**}	2.0083	0.2348^{*}	1.8184	0.1903	1.3013
a_{dJ}					-0.0726	-0.3162	0.0667	0.5544
a_{wJ}					0.2076	0.5957	0.2211	1.2746
a_{mJ}					0.0762	0.2313	0.1967	0.9362
$a_{\Delta J}$	-0.2181^{**}	-2.3698						
MZ-R^2	0.2284		0.2199		0.2214		0.2250	
MSE	0.8237		0.8549		0.8449		0.8447	
QLIKE	0.1900		0.1967		0.1941		0.1937	

[4] The Mincer Zarnowitz-R^2 is the R-squared of the true values of the dependent variable against its forecast values in the regression.

	h = 5							
a_0	0.5788***	4.0662	0.6047***	4.2318	0.5595***	4.1579	0.5801***	4.3122
a_d	0.1503**	2.2180	0.1040**	2.0137	0.1105*	1.7784	0.1553*	1.7049
a_w	0.2878***	2.6612	0.2777***	2.4228	0.4480***	2.7039	0.4069*	1.8499
a_m	0.2185*	1.8016	0.2265*	1.7832	0.1991*	1.7150	0.1893	0.8143
a_{dJ}					0.0711	0.4938	0.0448	0.5122
a_{wJ}					-0.4007	-1.5339	-0.0341	-0.1553
a_{mJ}					0.3509	0.8285	0.2917	0.9790
$a_{\Delta J}$	-0.2090**	-1.9957						
MZ-R^2	0.1255		0.1124		0.1208		0.1215	
MSE	0.9736		0.9901		0.9758		0.9773	
QLIKE	0.2278		0.2299		0.2280		0.2288	
	h = 22							
a_0	1.1778***	7.5183	1.1942***	7.5230	1.1834***	7.4257	1.1817***	7.4036
a_d	0.0865	1.2447	0.0612	1.0361	0.0943	1.1281	0.1099	1.2271
a_w	0.2117**	2.8376	0.2017**	2.3690	0.1998	1.2781	0.2234	1.1910
a_m	-0.0569	-0.4615	-0.0511	-0.4045	-0.0515	-0.2482	-0.1237	-0.5306
a_{dJ}					-0.0332	-0.3223	0.0044	0.0768
a_{wJ}					0.1666	0.6404	0.1245	0.7366
a_{mJ}					-0.0368	-0.0848	0.1163	0.3976
$a_{\Delta J}$	-0.1406**	-2.1892						
MZ-R^2	0.0305		0.0254		0.0292		0.0300	
MSE	1.0699		1.0770		1.0759		1.0748	
QLIKE	0.2904		0.2921		0.2920		0.2919	

Note: Table 2 presents the estimated coefficients as well as the t-statistics for each predictor in different models. The in-sample forecast performances are evaluated on the basis of the MZ-R^2, the MSE, and the QLIKE. *** denotes a significance level of 1%, ** denotes a significance level of 5%, and * denotes a significance level of 10%. h = 1 is the one-step-ahead forecast, h = 5 is the five-step ahead forecast, and h = 22 is the 22-step ahead forecast. The regression models are as follows:

HAR: $RV_t = a_0 + a_d \overline{RV}_{t,1} + a_w \overline{RV}_{t,5} + a_m \overline{RV}_{t,22} + u_t$

HAR-CJ: $RV_t = a_0 + a_d \overline{C}_{t,1} + a_w \overline{C}_{t,5} + a_m \overline{C}_{t,22} + a_{dJ} \overline{J}_{t,1} + a_{wJ} \overline{J}_{t,5} + a_{mJ} \overline{J}_{t,22} + u_t$

HAR-TCJ: $RV_t = a_0 + a_d \overline{TC}_{t,1} + a_w \overline{TC}_{t,5} + a_m \overline{TC}_{t,22} + a_{dJ} \overline{TJ}_{t,1} + a_{wJ} \overline{TJ}_{t,5} + a_{mJ} \overline{TJ}_{t,22} + u_t$

HAR-ΔJ: $RV_t = a_0 + a_{\Delta J} \overline{\Delta J}_{t,1} + a_d \overline{BPV}_{t,1} + a_w \overline{RV}_{t,5} + a_m \overline{RV}_{t,22} + u_t$

4.2 In-Sample Forecasts Using the HAR-type Models with Bayesian Approaches

We combine the HAR-type models with the DMA, the DMS, and the BMA approaches and use them to forecast the RV for both one-step and multi-step ahead forecasts, including the short-term forecast $h = 1$, the mid-term forecast $h = 5$, and the long-term forecast $h = 22$.

With Bayesian approaches, the predictors and the coefficients in the HAR-type models are no longer fixed but time-varying, and thus the disturbances from unknown shocks in the RV series can be eliminated. The total number of predictors (including constant) for the HAR-ΔJ model and the HAR model are 5 and 4 respectively, and for both the HAR-CJ model and the HAR-TCJ model, the total number of predictors is 7. Therefore, the number of sub-models equals $2^5 = 32$ for the HAR-ΔJ model, $2^4 = 16$ for the HAR model, and $2^7 = 128$ for both the HAR-CJ model and the HAR-TCJ model. We obtain the time-varying size of the predictors set in each model by calculating the weighted average of the number of predictors in each sub-model on the basis of the Bayesian weights:

$$E(Size_t^{DMA/BMA}) = \sum_{k=1}^{K} \pi_{t|t-1,k} Size_k$$

$$E(Size_t^{DMS}) = Size_{\hat{k}}\ ,\ \ \hat{k} : \pi_{t|t-1,\hat{k}} = \max\left\{\pi_{t|t-1,1}, \cdots, \pi_{t|t-1,l}, l = 1, \cdots, \mathrm{K}\right\},$$

where $Size_k$ is the number of predictors in each sub-model M_k and $\pi_{t|t-1,k}$ is the Bayesian probability that each sub-model is included in the model.

Table 3 shows the statistics of the time-varying sizes of predictor sets in various Bayesian models for different forecast steps. The time-varying sizes in the DMA and BMA models are the sum of the weighted sizes, while the size of a DMS model at *t* is equal to the size of the predictors in the sub-model with the highest inclusion probability. We divide the sample period into four sub-periods, and the length of each sub-period is around 1 year. We compute the means and standard deviations of the time-varying expected sizes in each sub-period, where the mean is the position indicator of sizes and the standard deviation scales the fluctuation of sizes. Table 3 clearly indicates that the sizes of the predictor sets in various Bayesian HAR-type models exhibit significant time-varying patterns. The average sizes of the predictors included in the Bayesian HAR-ΔJ models are around 3 to 4 for the short-term forecast, 2.5 to 3.5 for the mid-term forecast, and 2 to 3 for the long-term forecast. The average sizes of the predictor sets of the Bayesian HAR models are around 2.6 to 3.2 for the short-term forecast, 2 to 3.5 for the mid-term forecast, and 2 to 3.8 for the long-term forecast. For the Bayesian HAR-CJ models and HAR-TCJ models, there are about 3 to 5 predictors included in the short-term forecast models, 2.8 to 4.6 predictors included in the mid-term forecast models, and 2.8 to 4 predictors included in the long-term forecast models. In Table 3, we also present the maximum numbers of predictors allowed in the Bayesian models (or Max_N), which are also the numbers of predictors in the fixed-parameter HAR-type models. The differences between the means and the Max_N reflect the fact that the bad performance predictors with smaller weights are incorporated into the Bayesian HAR-type models over the forecast horizons. In conclusion, there are more predictors included in the short-term forecast model, while fewer predictors are included in the mid-term and long-term forecast models for all models except for the Bayesian HAR-CJ models. In addition, the standard deviations of sizes decrease over time in most Bayesian models, indicating less fluctuation of sizes during the later sample period.

Table 3 Statistics for the Time-Varying Sizes of Predictor Sets in Various HAR-type Models with Bayesian Approaches

	Start-April 2011		May 2011-April 2012		May 2012-April 2013		May 2013-End		Whole period
	mean	std	mean	std	mean	std	mean	std	Max_N
	h = 1								
DMA-HAR-ΔJ	3.1098	0.6534	3.7851	0.0555	3.6740	0.1358	3.5731	0.0714	5
DMS-HAR-ΔJ	3.1121	0.8737	4.0000	0.0000	3.3817	0.5346	3.2917	0.6177	5
BMA-HAR-ΔJ	3.4108	0.4214	3.7528	0.3188	3.9990	0.0014	3.9986	0.0026	5
DMA-HAR	2.7956	0.4577	3.2545	0.0529	3.2400	0.0543	3.0727	0.0790	4
DMS-HAR	2.6638	0.5160	3.0943	0.2922	3.0000	0.0000	2.9958	0.2327	4
BMA-HAR	2.9545	0.2154	2.1951	0.3617	2.0314	0.1004	2.9373	0.4310	4
DMA-HAR-CJ	3.8487	0.6981	4.2169	0.1705	4.3269	0.0995	4.1838	0.0913	7
DMS-HAR-CJ	3.9957	1.0107	4.4754	0.6102	4.1660	0.6858	3.2250	0.6641	7
BMA-HAR-CJ	3.9555	0.9139	4.1061	0.1629	3.6631	0.3911	4.0623	0.5200	7
DMA-HAR-TCJ	3.1333	0.9467	4.1936	0.1416	4.2023	0.1258	4.3734	0.1178	7
DMS-HAR-TCJ	2.9871	1.1314	4.4016	0.4985	3.3900	0.4878	3.8583	0.4339	7
BMA-HAR-TCJ	3.0633	1.0541	4.9614	0.0852	4.9998	0.0002	4.9999	0.0000	7
	h = 5								
DMA-HAR-ΔJ	2.7876	0.3500	2.6403	0.2019	3.2262	0.2267	3.5605	0.1712	5
DMS-HAR-ΔJ	2.7939	0.5352	2.2869	0.4523	2.8382	0.4007	3.7375	0.7373	5
BMA-HAR-ΔJ	2.9134	0.3313	2.8666	0.0614	2.4283	0.1875	2.8584	0.0681	5
DMA-HAR	3.1598	0.6846	2.3390	0.1609	2.6476	0.1679	3.0388	0.1321	4
DMS-HAR	3.2061	0.7760	2.0082	0.0902	2.3776	0.6589	3.5625	0.5517	4
BMA-HAR	3.3486	0.6111	2.7553	0.1318	2.1878	0.1584	2.6528	0.1453	4
DMA-HAR-CJ	4.1931	0.9669	3.5348	0.2831	4.4952	0.2535	4.4934	0.2757	7
DMS-HAR-CJ	4.2325	1.3680	3.0164	0.4524	4.6598	1.1703	3.9708	0.6916	7
BMA-HAR-CJ	4.1204	1.0241	3.1637	0.1612	3.0787	0.0196	3.1203	0.0162	7
DMA-HAR-TCJ	4.1399	0.7994	3.9285	0.2453	4.2177	0.1536	4.3158	0.2931	7
DMS-HAR-TCJ	4.0088	1.3011	3.9180	1.0290	3.6141	0.7543	3.4250	0.5725	7
BMA-HAR-TCJ	3.9192	0.9585	2.8500	0.1324	3.1530	0.1421	3.3271	0.2193	7
	h = 22								
DMA-HAR-ΔJ	2.2818	0.3628	2.8961	0.2152	3.0649	0.1168	3.0383	0.1005	5
DMS-HAR-ΔJ	2.2275	0.4724	2.7951	0.5112	2.6349	0.4815	2.2667	0.4422	5
BMA-HAR-ΔJ	2.7710	0.3662	2.9055	0.1840	2.8902	0.2304	2.6197	0.3083	5
DMA-HAR	2.0295	0.2054	2.1391	0.1298	2.5685	0.1059	2.6094	0.1409	4
DMS-HAR	1.9763	0.2656	2.0082	0.0902	2.0000	0.0000	2.2792	0.4843	4
BMA-HAR	2.3430	0.6771	2.0009	0.0022	3.4799	0.8375	3.8757	0.1999	4
DMA-HAR-CJ	3.8402	0.7883	3.4131	0.3989	4.0164	0.1916	3.9181	0.1521	7
DMS-HAR-CJ	3.4360	1.0969	3.1311	0.4782	2.8589	0.6731	2.9958	0.5283	7
BMA-HAR-CJ	3.2421	0.5252	3.0230	0.0503	3.0041	0.0375	2.9048	0.1878	7
DMA-HAR-TCJ	3.5825	0.5757	3.3828	0.3702	3.9691	0.1965	3.8585	0.3871	7
DMS-HAR-TCJ	3.4313	0.9075	3.2254	0.4984	3.1909	0.7491	2.9958	0.9552	7
BMA-HAR-TCJ	3.5485	0.6106	3.2632	0.5812	3.2430	0.5841	3.3768	0.2173	7

Note: Table 3 presents the mean and standard deviations (or std) of time-varying sizes in Bayesian HAR-type models for one-step (h = 1), five-step (h = 5), and 22-step (h = 22) ahead forecasts. Table 3 also presents the maximum number of predictors allowed in the model, or Max_N. The 4-year sample period is divided into four sub-periods: April 2010-April 2011, May 2011-April 2012, May 2012-April 2013, and May 2013-April 2014.

Further, we analyse the posterior inclusion probability (PIP) of different predictors in the DMA-HAR-type models and the BMA-HAR-type models. The posterior inclusion probabilities of the sub-models in DMS-HAR-type models are the same as those of the DMA-HAR-type models. For the DMS approach, we only select the sub-model with the highest probability at each time so that there is no need to compute the weighted PIP of each predictor.

The PIP of the i-th predictor within the DMA and BMA approaches is defined as

$$E(PIP_{i,t}) = \sum_{k=1}^{K} \pi_{t|t-1,k} * I(X_t^k \in sub\_M_k),$$

where $\pi_{t|t-1,k}$ is the Bayesian probability that the *k-th* sub-model is included in the model and *I( )* is the indication function whether the *i-th* predictor is included in the *k-th* sub-model (sub_M_k). A greater inclusion probability indicates a stronger explanatory power of this predictor: that is, it contains more information in the forecast. Based on Koop *et al.* (2012), the predictor is considered as a good one for a period if its inclusion probability is greater than 0.5.

Table 4 and Table 5 show the statistics of the time-varying PIP of predictors in the DMA-HAR-type and the BMA-HAR-type models, respectively. We present the means and standard deviations of posterior inclusion probabilities for each predictor for the short-term, mid-term, and long-term forecasts. In addition, we compute the proportion of high posterior inclusion probabilities (p(PIP>0.5)), which is equal to the accumulation of periods when the posterior inclusion probabilities are greater than 0.5 divided by the whole sample period. We consider a predictor to be a good one in the forecast model if either the mean of its PIP is greater than 0.5 or the proportion of its high posterior inclusion probabilities (PIP>0.5) is higher than 0.5, meaning that the periods when this predictor is a good one dominate the whole forecast period.

As shown in Table 4, when h = 1, the daily, weekly, and monthly volatility variables in four DMA-HAR-type models are the good predictors, with their proportions of high posterior inclusion probabilities being greater than 0.5. When h = 5, the daily and weekly volatility variables are the good predictors for all DMA-HAR-type models and the monthly jump variables are the good predictors for both the DMA-HAR-CJ model and the DMA-HAR-TCJ model. When h = 22, the monthly volatility variables are the good predictors for all DMA-HAR-type models and the daily jump variables are the good predictors for both the DMA-HAR-CJ model and the DMA-HAR-TCJ model.

As presented in Table 5, when h = 1, the daily and monthly volatility variables and the signed jump factor *ΔJ* are the good predictors for the BMA-HAR-ΔJ model and the monthly volatility is the only good predictor for the BMA-HAR model. The daily volatility variables and the daily and monthly jump variables are the good predictors for both the BMA-HAR-CJ model and the BMA-HAR-TCJ model, but the weekly volatility variable is

Table 4 Time-Varying Posterior Inclusion Probabilities of Predictors for the DMA-HAR-type Models

Models	DMA-HAR-ΔJ			DMA-HAR			DMA-HAR-CJ			DMA-HAR-TCJ		
	mean	std	p(PIP>0.5)	mean	std	p(PIP>0.5)	mean	std	p(PIP>0.5)	mean	std	p(PIP>0.5)
	h = 1											
PIP_d	**0.6339**	0.2725	**0.7202**	**0.6369**	0.2824	**0.6424**	**0.5573**	0.2213	**0.5636**	**0.5962**	0.2615	**0.5090**
PIP_w	**0.6051**	0.3176	**0.5838**	**0.7465**	0.2890	**0.8141**	**0.7551**	0.2591	**0.7525**	**0.7272**	0.2863	**0.8081**
PIP_m	**0.8732**	0.1570	**0.9455**	**0.7041**	0.2005	**0.8131**	**0.5389**	0.1686	**0.5232**	**0.5217**	0.1874	**0.5646**
PIP_{dJ}							0.4731	0.2439	0.3283	0.3883	0.1597	0.1283
PIP_{wJ}							0.3810	0.1628	0.1990	0.3416	0.1853	0.1455
PIP_{mJ}							0.4426	0.1496	0.2212	0.4202	0.1966	0.3687
$PIP_{\Delta J}$	0.4317	0.1841	0.3828									
	h = 5											
PIP_d	**0.5882**	0.2175	**0.6552**	**0.5128**	0.2010	0.3651	0.4388	0.2699	**0.5243**	**0.5600**	0.2341	**0.5923**
PIP_w	**0.7336**	0.1980	**0.9199**	**0.8875**	0.1292	**0.9899**	**0.7961**	0.1361	**0.9878**	**0.7562**	0.1651	**0.9848**
PIP_m	0.3603	0.3533	0.3692	0.3976	0.3549	0.3996	0.4062	0.2446	0.4047	0.3229	0.2231	0.2110
PIP_{dJ}							0.4164	0.1312	0.2292	0.3873	0.1542	0.2262
PIP_{wJ}							0.4931	0.1516	0.4919	**0.5438**	0.1849	0.4665
PIP_{mJ}							**0.6279**	0.1619	**0.6856**	**0.5802**	0.2225	**0.5862**
$PIP_{\Delta J}$	0.3878	0.1395	0.1653									
	h = 22											
PIP_d	0.3409	0.2029	0.0898	0.2473	0.2023	0.0805	0.3917	0.2448	0.2436	0.3366	0.2344	0.1455
PIP_w	0.3786	0.2969	0.3860	0.3603	0.3138	0.3715	0.4115	0.2295	0.3736	0.4268	0.2685	0.3251
PIP_m	**0.7061**	0.2675	**0.7740**	**0.7440**	0.2638	**0.7750**	**0.6628**	0.2230	**0.7730**	**0.6448**	0.2376	**0.7420**
PIP_{dJ}							**0.5552**	0.2403	0.4499	**0.5309**	0.2581	0.4469
PIP_{wJ}							0.3806	0.2097	0.2910	0.3525	0.2358	0.3127
PIP_{mJ}							0.3940	0.1942	0.2405	0.4088	0.2615	0.3849
$PIP_{\Delta J}$	0.4143	0.2664	0.2085									

Note: Table 4 presents the means and standard deviations of the time-varying posterior inclusion probabilities of different predictors in DMA-HAR-type models for one-step (h = 1), five-step (h = 5), and 22-step (h = 22) ahead forecasts. p(PIP>0.5) is the proportion of the posterior probabilities that are higher than 0.5 for each predictor. PIP_d, PIP_w, *and* PIP_m are the posterior inclusion probabilities of the daily, weekly, and monthly volatility variables, respectively; PIP_{dJ}, PIP_{wJ}, *and* PIP_{mJ} are the posterior inclusion probabilities of the daily, weekly, and monthly jump variables, respectively; and $PIP_{\Delta J}$ is the posterior inclusion probability of the signed jump variable in the DMA-HAR-ΔJ model. The good predictors are denoted in bold.

the good predictor only for the BMA-HAR-TCJ model. When h = 5, the daily and weekly volatility variables are the good predictors for both the BMA-HAR-ΔJ model and the BMA-HAR model. The weekly volatility variable and the monthly jump variable are the good predictors for the BMA-HAR-CJ model, while the weekly volatility variable and the weekly jump variable are the good predictors for the BMA-HAR-TCJ model. When h = 22, the monthly volatility variable and the signed jump factor *ΔJ* are the good predictors for the

BMA-HAR-ΔJ model and all variables in the BMA-HAR model are good predictors. The monthly volatility variables and the daily jump variables are the good predictors for both the BMA-HAR-CJ model and the BMA-HAR-TCJ model.

Table 5 Time-Varying Posterior Inclusion Probabilities of Predictors for the BMA-HAR-type Models

Models	BMA-HAR-ΔJ			BMA-HAR			BMA-HAR-CJ			BMA-HAR-TCJ		
	mean	std	p(PIP>0.5)	mean	std	p(PIP>0.5)	mean	std	p(PIP>0.5)	mean	std	p(PIP>0.5)
	h = 1											
PIP_d	**0.8518**	0.3222	**0.8465**	0.3760	0.4328	0.3838	**0.7314**	0.3724	**0.7667**	**0.9128**	0.2748	**0.9131**
PIP_w	0.1427	0.3166	0.1434	0.3242	0.4218	0.3192	0.3244	0.3495	0.2616	**0.8159**	0.3265	**0.8141**
PIP_m	**0.9968**	0.0375	**0.9960**	**0.8501**	0.2510	**0.8020**	0.4252	0.3822	0.4707	0.1173	0.2645	0.1242
PIP_{dJ}							**0.6555**	0.3600	**0.6010**	**0.8195**	0.3438	**0.8091**
PIP_{wJ}							0.2715	0.3774	0.2434	0.0449	0.0914	0.0040
PIP_{mJ}							**0.5747**	0.3827	**0.5283**	**0.8272**	0.3360	**0.8172**
$PIP_{\Delta J}$	**0.8102**	0.3275	**0.8222**									
	h = 5											
PIP_d	**0.6950**	0.2576	**0.7809**	**0.6300**	0.3196	**0.6988**	0.0584	0.1878	0.0588	0.2206	0.3214	0.1704
PIP_w	**0.9683**	0.1073	**0.9888**	**0.9910**	0.0877	**0.9899**	**0.9632**	0.1297	**0.9757**	**0.9571**	0.1045	**0.9899**
PIP_m	0.0341	0.1667	0.0284	0.1103	0.2938	0.1065	0.0789	0.2377	0.0771	0.0207	0.0971	0.0101
PIP_{dJ}							0.1035	0.1571	0.0406	0.0462	0.0750	0.0010
PIP_{wJ}							0.2196	0.3161	0.2028	**0.7869**	0.2775	**0.7718**
PIP_{mJ}							**0.9261**	0.1959	**0.9239**	0.2610	0.2791	0.1755
$PIP_{\Delta J}$	0.0725	0.0872	0.0071									
	h = 22											
PIP_d	0.2203	0.2663	0.1383	**0.5112**	0.4816	**0.5222**	0.1411	0.2847	0.1166	0.2863	0.3603	0.2776
PIP_w	0.2264	0.2816	0.1610	**0.5093**	0.4867	**0.5160**	0.1099	0.2204	0.0588	0.2114	0.3290	0.1754
PIP_m	**0.7330**	0.3627	**0.7575**	**0.9567**	0.1480	**0.9690**	**0.8270**	0.2948	**0.8338**	**0.7222**	0.3679	**0.7451**
PIP_{dJ}							**0.8362**	0.2713	**0.8927**	**0.7121**	0.3368	**0.7214**
PIP_{wJ}							0.0564	0.1043	0.0093	0.2767	0.3513	0.2570
PIP_{mJ}							0.0654	0.1315	0.0206	0.1432	0.1893	0.0609
$PIP_{\Delta J}$	**0.6084**	0.3856	**0.5934**									

Note: Table 5 presents the means and standard deviations (std) of the time-varying posterior inclusion probabilities for different predictors in BMA-HAR-type models for one-step (h = 1), five-step (h = 5), and 22-step (h = 22) ahead forecasts. p(PIP>0.5) is the proportion of the posterior probabilities that is greater than 0.5 for each predictor. PIP is the posterior inclusion probability. PIP_d, PIP_w, and PIP_m are the posterior inclusion probabilities of the daily, weekly, and monthly volatility variables, respectively; PIP_{dJ}, PIP_{wJ}, and PIP_{mJ} are the posterior inclusion probabilities of the daily, weekly, and monthly jump variables, respectively; and $PIP_{\Delta J}$ is the posterior inclusion probability of the signed jump variable in BMA-HAR-ΔJ model. The good predictors are denoted in bold.

Table 6 Time-Varying Coefficients for Various HAR-type Models with Bayesian Approaches

Models		HAR-ΔJ		HAR		HAR-CJ		HAR-TCJ	
		DMA	BMA	DMA	BMA	DMA	BMA	DMA	BMA
	h = 1								
a_d	mean	0.1281*	0.3253**	0.0732*	0.0974*	0.0916*	0.2470**	0.1113*	0.1723**
	std	0.1459	0.1321	0.1197	0.1502	0.1298	0.1586	0.1423	0.1177
a_w	mean	0.2855**	0.0533	0.3573**	0.0995*	0.4250**	0.1287*	0.3789**	0.3785**
	std	0.2283	0.1589	0.1974	0.1525	0.2577	0.2492	0.2666	0.2357
a_m	mean	0.1780	0.1497	0.0603	-0.0345	0.1782*	0.1196	0.3544	0.1552
	std	0.4280	0.4469	0.4119	0.4561	0.2599	0.2812	0.8745	0.8530
a_{dJ}	mean					-0.0302	-0.1095**	0.0124	0.1230**
	std					0.0636	0.1053	0.0467	0.0613
a_{wJ}	mean					-0.0851	0.0446	0.0912	0.0004
	std					0.4296	0.4328	0.2317	0.1355
a_{mJ}	mean					-0.0260	0.2350	-0.1227	0.1104
	std					1.5330	1.5225	2.2261	2.1849
$a_{\Delta J}$	mean	-0.0578**	-0.1959**						
	std	0.0562	0.0802						
	h = 5								
a_d	mean	0.1030*	0.0113*	0.0234	0.0218	0.0584*	-0.0135	0.0372	-0.0455
	std	0.1031	0.0168	0.0612	0.0608	0.1053	0.0585	0.1566	0.1171
a_w	mean	0.2660**	0.5263**	0.3741**	0.5477**	0.4296**	0.7085**	0.4756*	0.8006**
	std	0.2338	0.2163	0.2317	0.1686	0.3285	0.2776	0.4918	0.3961
a_m	mean	0.1780	0.0719	0.1295	0.0384	0.1480	0.0577	0.1297	0.0540
	std	0.3786	0.3830	0.2776	0.2409	0.3469	0.3437	0.5826	0.5553
a_{dJ}	mean					-0.0112	0.0211	0.0282*	0.0037
	std					0.1281	0.0911	0.0406	0.0083
a_{wJ}	mean					-0.3249	-0.2642	-0.0641	0.0274
	std					0.7864	0.7939	0.4145	0.4423
a_{mJ}	mean					0.7075*	0.8815*	0.4367*	0.1563
	std					1.2427	1.0722	0.7943	0.6121
$a_{\Delta J}$	mean	0.0108	0.0128*						
	std	0.0765	0.0252						
	h = 22								
a_d	mean	0.0109	-0.0046	0.0026	-0.0074	-0.0045	-0.0126	0.0046	-0.0182
	std	0.0494	0.0550	0.0662	0.0642	0.0905	0.0758	0.0793	0.0720
a_w	mean	0.0205	-0.0389	0.0324	-0.0234	-0.0231	-0.0370	-0.0378	-0.0430
	std	0.1245	0.1260	0.0951	0.1777	0.1889	0.1376	0.3601	0.3022
a_m	mean	0.1165*	0.1786*	0.0801	0.0276	0.0737	0.0617	0.1277	0.1522
	std	0.2011	0.2254	0.2836	0.2766	0.3725	0.3938	0.3576	0.4410
a_{dJ}	mean					0.0058	0.0994**	-0.0200	-0.0417*
	std					0.1205	0.0643	0.0465	0.0448
a_{wJ}	mean					0.0655	0.0050	0.0491	0.1103
	std					0.4313	0.1975	0.2325	0.2336
a_{mJ}	mean					-0.0100	-0.1027	0.1257	-0.0359
	std					1.0456	0.5447	0.5390	0.2833
$a_{\Delta J}$	mean	-0.0305*	-0.0763**						
	std	0.0549	0.0721						

Note: Table 6 shows the means and standard deviations (std) of the time-varying coefficients for various HAR-type models with Bayesian approaches for one-step (h = 1), five-step (h = 5), and 22-step (h = 22) ahead forecasts. a_d, a_w, and a_m are the coefficients for the daily, weekly, and monthly volatility variables; a_{dJ}, a_{wJ}, and a_{mJ} are the coefficients for the daily, weekly, and monthly jump variables; and $\alpha_{\Delta J}$ is the coefficient for the signed jump factor in the HAR-ΔJ model. We mark the predictors with a |mean/std| greater than 1 with two stars and the predictors with a |mean/std| greater than 0.5 with one star.

Table 6 shows the average means and standard deviations of expected coefficients over the sample period. Within the DMS approach, only the sub-model with the highest PIP is employed in the forecast model for each period, which indicates that the coefficients for predictors are time-discrete rather than time-continuous.[5] Therefore, we only list the statistics of time-varying coefficients in the DMA-HAR-type models and the BMA-HAR-type models. The coefficient of each predictor within the DMA and BMA approaches are computed by

$$E(\hat{\beta}_t^{DMA/BMA}) = \sum_{k=1}^{K} \pi_{t+1|t,k} \hat{\beta}_t^{(k)} .$$

As shown in Table 6, the means and standard deviations suggest significant time-varying characteristics of coefficients in the HAR models with Bayesian approaches. The absolute ratio of the posterior mean against the standard deviation ($|mean/std|$) of a coefficient in the Bayesian models is used as a criterion of stable predictability. A high $|mean/std|$ means the stable predictability of a predictor over time and vice versa. We mark the predictors with a $|mean/std|$ greater than 1 with two stars and the predictors with a $|mean/std|$ greater than 0.5 with one star. The results in Table 6 show that $|mean/std|$ decreases with the increase in forecast steps, indicating the weaker predictability of predictors in the longer term forecast. Table 6 also shows that when h = 1, the daily and weekly volatility variables, the daily jump variable, and the signed jump variable ΔJ have stable predictability. When h = 5, the weekly volatility variables and the monthly jump variables have stable predictability. When h = 22, the monthly volatility variable, the daily jump variables, and the signed jump variable ΔJ have stable predictability. Moreover, the coefficients in the Bayesian HAR-type models are smaller than those in the fixed parameter models, indicating that all the predictors within the DMA and BMA approaches have time-varying predictability so that they can be incorporated into the forecast models on the basis of their forecast performances over time.

Table 7 presents the results of the in-sample forecast performances for 12 HAR-type models with Bayesian approaches based on three types of statistics: the MZ-R^2, the MSE loss function, and the QLIKE loss function. The in-sample results are based on the last 800 forecast samples (about 80% of the sample size) excluding some outliers in the initial iterations of the Kalman filter. As shown in Table 7, all types of statistics suggest that compared with the other Bayesian models, the DMS-HAR-ΔJ model, the DMS-HAR-CJ model, and the DMS-HAR-TCJ model have the superior performances for the short-term forecast and the mid-term forecast. Among these models, the DMS-HAR-CJ model

---

[5] Time-discrete means that the predictors do not have the posterior weighted average coefficients for all periods (the DMS-HAR-type models). On the contrary, time-continuous means that all the predictors have the posterior weighted average coefficients at each period *t* (the DMA-HAR-type and the BMA-HAR-type models).

performs best for both the short-term forecast and the mid-term forecast. For the long-term forecast, all the evaluation approaches suggest that the BMA-HAR-TCJ model performs best. In addition, the HAR-type models with Bayesian approaches perform better than their original HAR-type models for the in-sample forecast.

Table 7 In-Sample Forecast Comparisons among Various Forecast Models

	h = 1			h = 5			h = 22		
	MZ-R^2	MSE	QLIKE	MZ-R^2	MSE	QLIKE	MZ-R^2	MSE	QLIKE
	Bayesian models								
DMA-HAR-ΔJ	0.2469	0.6981	0.1875	0.1447	0.7997	0.2214	0.0346	0.9357	0.2888
DMS-HAR-ΔJ	**0.2510**	**0.6940**	**0.1843**	**0.1519**	**0.7926**	**0.2190**	0.0363	0.9357	0.2886
BMA-HAR-ΔJ	0.2452	0.7074	0.1903	0.1257	0.8274	0.2291	**0.0496**	**0.9132**	0.2781
DMA-HAR	0.2429	0.7012	0.1891	0.1368	0.8081	0.2243	0.0312	0.9324	0.2845
DMS-HAR	0.2461	0.6983	0.1886	0.1411	0.8034	0.2212	0.0343	0.9291	0.2813
BMA-HAR	0.1703	0.8006	0.2215	0.1226	0.8296	0.2300	0.0200	0.9519	0.2868
DMA-HAR-CJ	0.2555	0.6907	0.1852	0.1441	0.7993	0.2221	0.0373	0.9275	0.2824
DMS-HAR-CJ	**0.2686**	**0.6784**	**0.1787**	**0.1548**	**0.7889**	**0.2187**	**0.0432**	**0.9093**	**0.2666**
BMA-HAR-CJ	0.2340	0.7121	0.1940	0.1363	0.8103	0.2224	0.0199	0.9335	**0.2772**
DMA-HAR-TCJ	0.2547	0.6911	0.1873	0.1413	0.8026	0.2241	0.0338	0.9327	0.2828
DMS-HAR-TCJ	**0.2668**	**0.6802**	**0.1832**	**0.1515**	**0.7938**	**0.2195**	0.0421	0.9244	0.2775
BMA-HAR-TCJ	0.2615	0.6865	0.1866	0.1281	0.8144	0.2250	**0.0540**	**0.9068**	**0.2655**
	Benchmark models								
HAR-ΔJ	0.2284	0.8237	0.1900	0.1255	0.9736	0.2278	0.0305	1.0699	0.2904
HAR	0.2199	0.8549	0.1967	0.1124	0.9901	0.2299	0.0254	1.0770	0.2921
HAR-CJ	0.2214	0.8449	0.1941	0.1208	0.9758	0.2280	0.0292	1.0759	0.2920
HAR-TCJ	0.2250	0.8447	0.1937	0.1215	0.9773	0.2288	0.0300	1.0748	0.2919

Note: Table 7 presents the in-sample forecast comparisons among the various forecast models for the one-step (h = 1), five-step (h = 5), and 22-step (h = 22) ahead forecasts. The in-sample forecast performances are evaluated on the basis of the MZ-R^2, the MSE loss function, the QLIKE loss function, and the sum of log predicted likelihood. The HAR-type models are used as the benchmark models.

4.3 Out-of-Sample Forecasts

We divide the whole sample into two parts: the in-sample part that contains the first 491 observations from 16 April 2010 to 23 April 2012, and the out-of-sample part that contains the last 500 observations from 24 April 2012 to 21 May 2014. We employ the recursive forecast to obtain the short-term ($h = 1$), mid-term ($h = 5$), and long-term ($h = 22$) out-of-sample forecast values corresponding to one-day, one-week, and one-month ahead forecasts. Patton's (2011) loss functions are utilised to compare the forecast performances of various competing models. We employ four loss functions: the MSE and the QLIKE loss functions when b = 0 and b = -2 respectively and the homogeneous robust loss function and the positive robust loss function when $b = -1$ and $b = 1$ respectively. We obtain 12 candidate models with time-varying parameters and variable predictor sets by combining the

HAR-type models with the three Bayesian approaches, and the conventional HAR-type models are used as the benchmark models. The results are reported in Table 8.

Table 8 List of the Forecast Models Ranked by the Loss Functions

	h = 1		h = 5		h = 22		h = 1		h = 5		h = 22	
Models	value	rank	value	rank	value	rank	value	rank	value	rank	value	rank
	MSE loss function (b = 0)						QLIKE loss function (b = -2)					
DMA-HAR-ΔJ	**1.1244**	**2**	1.0724	5	1.3564	9	**0.2500**	**2**	0.2379	4	0.3448	11
DMS-HAR-ΔJ	**1.1246**	**3**	**1.0640**	**2**	1.3537	6	**0.2502**	**3**	**0.2355**	**2**	0.3428	8
BMA-HAR-ΔJ	1.3332	11	1.1158	11	**1.3231**	**3**	0.3296	12	0.2491	10	**0.3263**	**3**
DMA-HAR	1.1279	4	1.0851	8	1.3552	7	0.2506	4	0.2423	9	0.3439	12
DMS-HAR	1.1288	5	1.0775	7	1.3557	8	0.2510	5	0.2381	5	0.3425	7
BMA-HAR	1.2655	10	1.1184	12	1.3713	12	0.3170	10	0.2507	11	0.3405	6
DMA-HAR-CJ	1.1532	6	1.0701	4	1.3526	5	0.2588	6	0.2387	6	0.3432	9
DMS-HAR-CJ	1.1540	7	**1.0571**	**1**	**1.3228**	**2**	0.2596	7	**0.2349**	**1**	**0.3155**	**2**
BMA-HAR-CJ	1.3606	12	1.0862	9	1.3593	11	0.3261	11	0.2391	7	0.3314	4
DMA-HAR-TCJ	1.1845	8	1.0727	6	1.3581	10	0.2754	8	0.2388	7	0.3445	10
DMS-HAR-TCJ	**0.9673**	**1**	**1.0643**	**3**	1.3458	4	**0.2019**	**1**	**0.2364**	**3**	0.3378	5
BMA-HAR-TCJ	1.1845	9	1.0888	10	**1.2952**	**1**	0.2754	9	0.2412	8	**0.3034**	**1**
HAR-ΔJ	1.6085	14	1.8982	14	2.3344	14	0.3612	14	0.3932	13	0.4229	14
HAR	1.5937	13	1.8810	13	2.2556	13	0.3565	13	0.3936	14	0.4225	13
HAR-CJ	1.6689	16	1.9783	16	2.3886	16	0.3645	15	0.3986	15	0.4290	15
HAR-TCJ	1.6650	15	1.9746	15	2.3882	15	0.3650	16	0.3997	16	0.4304	16
	Homogeneous robust loss function (b = -1)						Positive robust loss function (b = 1)					
DMA-HAR-ΔJ	**0.4206**	**2**	0.3937	5	0.5525	10	**5.5482**	**2**	5.4872	6	6.2467	8
DMS-HAR-ΔJ	0.4207	4	**0.3895**	**3**	0.5502	7	**5.5486**	**3**	**5.4670**	**2**	6.2430	5
BMA-HAR-ΔJ	0.5744	11	0.4131	11	**0.5253**	**3**	5.8469	11	5.6202	12	**6.2120**	**2**
DMA-HAR	**0.4206**	**3**	0.4011	10	0.5516	9	5.5601	4	5.5095	8	6.2449	7
DMS-HAR	0.4213	5	0.3955	8	0.5507	8	5.5614	5	5.5029	7	6.2484	9
BMA-HAR	0.5151	10	0.4155	12	0.5559	12	5.7687	10	5.6147	11	6.2807	12
DMA-HAR-CJ	0.4354	6	0.3935	4	0.5492	6	5.6057	8	5.4822	4	6.2436	6
DMS-HAR-CJ	0.4362	7	**0.3882**	**1**	**0.5206**	**2**	5.6066	9	**5.4490**	**1**	**6.2149**	**3**
BMA-HAR-CJ	0.6012	12	0.3965	6	0.5450	5	5.8749	12	5.5674	10	6.2690	11
DMA-HAR-TCJ	0.4578	8	0.3944	7	0.5529	11	5.6501	6	5.4865	5	6.2514	10
DMS-HAR-TCJ	**0.3374**	**1**	**0.3892**	**2**	0.5440	4	**5.2154**	**1**	**5.4694**	**3**	6.2338	4
BMA-HAR-TCJ	0.4579	9	0.4001	9	**0.5020**	**1**	5.6502	7	5.5513	9	**6.1745**	**1**
HAR-ΔJ	0.6370	14	0.7299	14	0.8265	13	6.7399	13	7.7071	13	9.3294	13
HAR	0.6320	13	0.7254	13	0.8346	14	6.9390	14	7.9349	14	9.9689	15
HAR-CJ	0.6530	15	0.7517	15	0.8542	15	7.1207	15	8.1757	16	10.0143	16
HAR-TCJ	0.6531	16	0.7530	16	0.8568	16	7.0935	16	8.1297	15	9.9615	14

Note: Table 8 compares the out-of-sample performances among the various forecast models for one-step (h = 1), five-step (h = 5), and 22-step (h = 22) ahead forecasts on the basis of the MSE, the QLIKE, the homogeneous robust loss function, and the positive robust loss function. The top three ranked models are in bold.

Table 8 shows a list of forecast models ranked by Patton's (2011) loss functions; the list suggests that the HAR-type models with the Bayesian approaches perform better than the corresponding HAR models. According to the results of the four loss functions, the forecast performances of the combination models are 20 to 50 per cent higher than those of the corresponding HAR models. When h = 1, the DMS-HAR-TCJ model, the DMA-HAR-ΔJ model, and the DMS-HAR-ΔJ model are ranked the first three models among the competing models on the basis of the MSE, the QLIKE, and the positive loss functions. When h = 5, all the loss functions suggest that the first three models are the DMS-HAR-CJ model, the DMS-HAR-ΔJ model, and the DMS-HAR-TCJ. On the basis of all the loss functions, when h = 22, the first three models are the BMA-HAR-TCJ, the DMS-HAR-CJ, and the BMA-HAR-ΔJ. In general, all the loss functions suggest that the DMS-HAR-TCJ model, the DMS-HAR-CJ model, and the BMA-HAR-TCJ model perform best in the short term, the midterm, and the long term, respectively, improving their benchmark HAR-type models by 41.9%, 46.6%, and 45.7% respectively on the basis of the MSE and 44.7%, 41.1%, and 29.5% respectively on the basis of the QLIKE.

We also compare all the forecast models on the basis of the MCS test of Hansen *et al.* (2011). The MSE and QLIKE are used as the loss functions of the MCS test. The number of bootstrap samples used to obtain the p-value that the null is rejected is set at 10,000, where the greater the p-value of a forecast model, the higher the probability that it belongs to the optimal forecast model set. The results are summarised in Table 9.

As shown in Table 9, for the short-term forecast models (h = 1), all the Bayesian models are included in the MCS at the 10% confidence level according to T_R statistics and only the DMS-HAR-TCJ model, the DMA-HAR-ΔJ model, and the DMA-HAR-ΔJ model are included in the MCS at the 10% confidence level within both the MSE and QLIKE loss functions. For the mid-term forecast models (h = 5), all the Bayesian models are included in the MCS at the 10% confidence level on the basis of both T_R and T_{SQ} within the two loss functions. Finally, for the long-term forecast models (h = 22), all the forecast models are included in the MCS at the 10% confidence level. The best models for the short-term forecast and the long-term forecast are the DMS-HAR-TCJ model and the BMA-HAR-TCJ model, respectively. For the mid-term forecast, the models that perform best are the DMS-HAR-CJ model based on the MSE loss function and the DMS-HAR-TCJ model based on the QLIKE loss function.

The DMS approach improves the forecast performances of the HAR-type models in both the short term and midterm to the greatest extent, while the BMA approach improves the forecast performances of the HAR-type models in the long term for both in-sample and out-of-sample forecasts, as shown in tables 7 through 9. Particularly, the DMS-HAR-CJ model and the BMA-HAR-TCJ model perform best for both the in-sample and out-of-sample forecasts in the midterm and long term, respectively. The DMS-HAR-CJ model performs best for the in-sample forecast, while the DMS-HAR-TCJ model performs best for the out-of-sample forecast in the short term.

Table 9 MCS Results of All Forecast Models Based on the MSE and QLIKE Loss Functions

		h = 1		h = 5		h = 22	
Group	Models	$Pval_R$	$Pval_{SQ}$	$Pval_R$	$Pval_{SQ}$	$Pval_R$	$Pval_{SQ}$
		MSE loss function					
1	DMA-HAR-ΔJ	0.305**	0.222*	0.443**	0.333**	0.380**	0.495**
2	DMS-HAR-ΔJ	0.305**	0.222*	0.785**	0.763**	0.648**	0.542**
3	BMA-HAR-ΔJ	0.240*	0.044	0.365**	0.135$^{*\backslash}$	0.653**	0.628**
4	DMA-HAR	0.305**	0.2220*	0.365**	0.179*	0.380**	0.495**
5	DMS-HAR	0.305**	0.2220*	0.443**	0.275**	0.653**	0.528**
6	BMA-HAR	0.240*	0.0560	0.365**	0.313**	0.365**	0.495**
7	DMA-HAR-CJ	0.264**	0.0960	0.443**	0.296**	0.380**	0.628**
8	DMS-HAR-CJ	0.264**	0.0750	**1.000****	**1.000****	0.653**	0.628**
9	BMA-HAR-CJ	0.155*	0.0380	0.443**	0.110*	0.380**	0.495**
10	DMA-HAR-TCJ	0.264**	0.0820	0.443**	0.313**	0.365**	0.495**
11	DMS-HAR-TCJ	**1.000****	**1.0000****	0.785**	0.763**	0.653**	0.554**
12	BMA-HAR-TCJ	0.264**	0.1000*	0.443**	0.257**	**1.000****	**1.000****
13	HAR-ΔJ	0.000	0.0120	0.000	0.000	0.294**	0.220**
14	HAR	0.000	0.0160	0.000	0.000	0.294**	0.079
15	HAR-CJ	0.000	0.0200	0.000	0.004	0.294**	0.132*
16	HAR-TCJ	0.000	0.0290	0.000	0.000	0.275**	0.0980
		QLIKE loss function					
1	DMA-HAR-ΔJ	0.235*	0.141*	0.879**	0.844**	0.341**	0.203*
2	DMS-HAR-ΔJ	0.235*	0.141*	0.923**	0.913**	0.341**	0.377**
3	BMA-HAR-ΔJ	0.108*	0.04	0.293**	0.158*	0.341**	0.387**
4	DMA-HAR	0.235*	0.094	0.502**	0.158*	0.443**	0.431**
5	DMS-HAR	0.235*	0.094	0.879**	0.844**	0.341**	0.258**
6	BMA-HAR	0.164*	0.054	0.188*	0.158*	0.443**	0.431**
7	DMA-HAR-CJ	0.235*	0.065	0.728**	0.413**	0.341**	0.377**
8	DMS-HAR-CJ	0.235*	0.058	0.923**	0.913**	0.621**	0.621**
9	BMA-HAR-CJ	0.108*	0.035	0.879**	0.844**	0.443**	0.431**
10	DMA-HAR-TCJ	0.235*	0.08	0.731**	0.413**	0.341**	0.387**
11	DMS-HAR-TCJ	**1.000****	**1.000***	**1.000****	**1.000***	0.443**	0.431**
12	BMA-HAR-TCJ	0.235*	0.094	0.728*	0.413**	**1.000****	**1.000****
13	HAR-ΔJ	0.033	0.028	0.026	0.043	0.341**	0.373**
14	HAR	0.031	0.022	0.008	0.038	0.341**	0.364**
15	HAR-CJ	0.031	0.016	0.003	0.005	0.341**	0.344**
16	HAR-TCJ	0.031	0.016	0.004	0.005	0.341**	0.329**

Note: Table 9 compares the out-of-sample performances among various forecast models for the one-step (h = 1), five-step (h = 5), and 22-step (h = 22) ahead forecasts on the basis of the MCS results. In addition, p-values marked with two stars are in $\hat{M}^*_{0.75}$, and p-values marked with one star are in $\hat{M}^*_{0.90}$. Note that $\hat{M}^*_{0.75} \subset \hat{M}^*_{0.90}$.

V. Conclusion

The conventional forecast models normally assume constant parameters and invariable predictor sets. We investigate the RV forecast models with time-varying parameters and variable predictors by combining the four representative HAR-type models with the three Bayesian approaches (DMA, DMS, and BMA) and by employing the high-frequency data of the CSI 300 futures from 16 April 2010 to 21 May 2014. We analyse the time-varying weighted average sizes of each Bayesian model as well as the time-varying posterior inclusion probabilities and weighted average coefficients of each predictor for different forecast steps. We compare the in-sample forecast performances of all the models on the basis of the MZ-R^2, the MSE, and the QLIKE loss function and the out-of-sample forecast performances of all the models on the basis of the loss functions of Patton (2011) and the MCS of Hansen *et al.* (2011).

The forecast results show the time-varying sizes of predictor sets in different Bayesian models for the different forecast steps. The daily, weekly, and monthly volatility variables perform well depending on the type of HAR models based on the results of their posterior inclusion probabilities. Compared with their benchmark models, the HAR-type models with Bayesian approaches perform better for both in-sample and out-of-sample forecasts. On the basis of most of the types of statistics used in the study, for in-sample forecasts, the DMS-HAR-CJ model performs best for both short-term and mid-term forecasts while the BMA-HAR-TCJ model performs best for long-term forecasts. On the basis of all the loss functions and the MCS results, for out-of-sample forecasts, the DMS-HAR-TCJ model, the DMS-HAR-CJ model, and the BMA-HAR-TCJ model perform best in the short term, midterm, and long term, respectively. Particularly, the DMS-HAR-CJ model and the BMA-HAR-TCJ model perform best in the midterm and long term respectively for both in-sample and out-of-sample forecasts. In general, the DMS approach improves the forecast performances of the HAR-type models for both short-term and mid-term forecasts to the greatest extent, while the BMA approach improves forecast performance in the long term for both in-sample and out-of-sample forecasts.

References

Ahoniemi, K. and Lanne, M. (2013), ‘Overnight stock returns and realized volatility’, *International Journal of Forecasting* 29 (4): 592-604.

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’, *The Review of Economics and Statistics* 89 (4): 701-720.

Bandi, F. M. and Renò, R. (2012), ‘Time-varying leverage effects’, *Journal of Econometrics* 169 (1): 94-113.

Barndorff-Nielsen, O. E., Kinnebrock, S., and Shephard, N. (2010), ‘Measuring downside risk – realised semivariance’, in Bollerslev, T., Russell, J., and Watson, M. (eds.), V*olatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, Oxford University Press, pp. 117-136.

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.

Barndorff-Nielsen, O. E. and Shephard, N. (2006), ‘Econometrics of testing for jumps in financial economics using bipower variation’, *Journal of Financial Econometrics* 4 (1): 1-30.

Beltratti, A. and Morana, C. (2005), ‘Statistical benefits of value-at-risk with long memory’, *Journal of Risk* 7 (4): 21-45.

Choi, K., Yu, W. C., and Zivot, E. (2010), ‘Long memory versus structural breaks in modeling and forecasting realized volatility’, *Journal of International Money and Finance* 29 (5): 857-875.

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

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 and Economic Statistics* 30 (3): 368-380.

Engle, R. F. and Gallo, G. M. (2006), ‘A multiple indicators model for volatility using intra-daily data’, *Journal of Econometrics* 131 (1): 3-27.

Engle, R. F., Ghysels, E., and Sohn, B. (2008), ‘On the economic sources of stock market volatility’, *AFA 2008 New Orleans Meetings Paper*, 31 August 2008.

Goldman, E., Nam, J., Tsurumi, H., and Wang, J. (2013), ‘Regimes and long memory in realized volatility’, *Studies in Nonlinear Dynamics and Econometrics* 17 (5): 521-549.

Hansen, P. R. and Lunde, A. (2006), ‘Realized variance and market microstructure noise’, *Journal of Business and Economic Statistics* 24 (2): 127-161.

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

Koop, G. and Korobilis, D. (2012), 'Forecasting Inflation Using Dynamic Model Averaging', *International Economic Review* 53 (3): 867-886.

Koopman, S. J., Jungbacker, B., and Hol, E. (2005), 'Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements', *Journal of Empirical Finance* 12 (3): 445-475.

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

Martens, M. (2002), 'Measuring and forecasting S&P 500 index-futures volatility using high-frequency data', *Journal of Futures Markets* 22 (6): 497-518.

Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Picteta, O. V., and von Weizsäcker, J. E. (1997), 'Volatilities of different time resolutions – Analyzing the dynamics of market components', *Journal of Empirical Finance* 4 (2-3): 213-239.

Patton, A. J. (2011), 'Volatility forecast comparison using imperfect volatility proxies', *Journal of Econometrics* 160 (1): 246-256.

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.

Politis, D. N. and Romano, J. P. (1994), 'Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions', *The Annals of Statistics* 22 (4): 2031-2050.

Raftery, A. E., Kárný, 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.

Tsiakas, I. (2008), 'Overnight information and stochastic volatility: A study of European and US stock exchanges', *Journal of Banking and Finance* 32 (2): 251-268.

Wright, J. H. (2008), 'Bayesian model averaging and exchange rate forecasts', *Journal of Econometrics* 146 (2): 329-341.

Xu, J. and Perron, P. (2014), 'Forecasting return volatility: Level shifts with varying jump probability and mean reversion', *International Journal of Forecasting* 30 (3): 449-463.

Appendix A Data Generation Process and Partial Variance

The data generation process for price p_t is a stochastic process with continuous components and jump components:

$$p_t = \int_t^{t+1} b_s ds + \int_t^{t+1} \sigma_s dW_s + \sum_{j=1}^{N} c_{jt} , \tag{A1}$$

where b_s is a locally bound and predictable drift, σ_s is a diffusive càdlàg process, W_s is a standard Brownian motion, c_{jt} is a non-zero random value, and N is the number of jumps.

The non-parameter recursive filter $\hat{V}_j$ in the partial variance is estimated by iterating in z:

$$\hat{V}_j^z = \frac{\sum_j K(\frac{j}{L}) r_j^2 I(\{r_g^2 \hat{V}_j^{z-1}\}}{\sum_j K(\frac{j}{L}) I(\{r_j^2 \le c_g^2 \hat{V}_j^{z-1}\}} , \tag{A2}$$

where $j = \{x \in Z : x = [-L, L] \wedge x \ne [-1,1]\}$ and $z = 1, 2, \cdots$, and $K(.)$ is a Gaussian kernel $K(y) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{y^2}{2})$.

Appendix B Estimation of DMA

The Kalman filter is employed in DMA to update the parameters in each sub-model of the forecast model.

For a one-model case, the model can be simplified as follows:

$$RV_t \sim N(X_t \beta_t, \mathrm{H}_t) \tag{B1}$$

$$\beta_t \sim N(\beta_{t-1}, Q_t) \tag{B2}$$

The inference of the model parameters is achieved on the basis of the Kalman filter. Assume that

$$\beta_{t-1} \mid RV^{t-1} \sim N(\hat{\beta}_{t-1}, \Sigma_{t-1|t-1}) \tag{B3}$$

As $\Sigma_{t|t-1} = \Sigma_{t-1|t-1} + Q_t$,

$$\beta_t \mid RV^{t-1} \sim N(\hat{\beta}_{t-1}, \Sigma_{t|t-1}) \tag{B4}$$

The updating equation for $\hat{\beta}_t$ is

$$\hat{\beta}_t = \hat{\beta}_{t-1} + \Sigma_{t|t-1} X_t (\mathrm{H}_t + X_t \Sigma_{t|t-1} X_t')^{-1} (RV_t - X_t \hat{\beta}_{t-1}) \tag{B5}$$

and

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1} X_t (\mathrm{H}_t + X_t \Sigma_{t|t-1} X_t')^{-1} X_t \Sigma_{t|t-1} \tag{B6}$$

The recursive forecast of RV_t can be conducted on the basis of the predictive density:

$$RV_t \mid RV^{t-1} \sim N(X_t\hat{\beta}_{t-1}, \mathrm{H}_t + X_t\Sigma_{t|t-1}X_t{}') \quad \text{(B7)}$$

It follows that

$$\tilde{\mathrm{H}}_t = \frac{1}{t}\sum_{j=1}^{t}\left[(RV_j - X_j\hat{\beta}_{j-1}) - X_j\,\Sigma_{j|j-1}\,X_j{}'\right], \quad \text{(B8)}$$

where $\tilde{\mathrm{H}}_t$ is a consistent estimator of H_t, and $\tilde{\mathrm{H}}_t \to \mathrm{H}_t$, as $t \to \infty$. In the multi-model case, $\tilde{\mathrm{H}}_t$ in each sub-model is denoted as $\tilde{\mathrm{H}}_t^{(k)}$:

$$\tilde{\mathrm{H}}_t^{(k)} = \frac{1}{t}\sum_{j=1}^{t}\left[(RV_j - X_j^{(k)}\hat{\beta}_{j-1}^{(k)}) - X_j^{(k)}\,\Sigma_{j|j-1}^{(k)}\,X_j^{(k)}{}'\right] \quad \text{(B9)}$$

On the basis of Raftery *et al*. (2010), we can avoid the rare possibility of $\tilde{\mathrm{H}}_t^{(k)} < 0$ by substituting $\mathrm{H}_t^{(k)}$ for $\hat{\mathrm{H}}_t^{(k)}$ in (B7):

$$\hat{\mathrm{H}}_t^{(k)} = \begin{cases} \tilde{\mathrm{H}}_t^{(k)}, & \text{if } \tilde{\mathrm{H}}_t^{(k)} > 0 \\ \hat{\mathrm{H}}_t^{(k)}, & \text{otherwise} \end{cases} \quad \text{(B10)}$$

The conversion of a multi-model case to a one-model case is based on the inclusion probability of each sub-model. Suppose that we know the conditional distribution of the state at time (t - 1) given the data up to that time: the probability distribution of the underlying state (L_t, β_t) can be written as

$$p(L_{t-1}, \beta_{t-1} \mid RV^{t-1}) = \sum_{k=1}^{K} p(\beta_{t-1}^{(k)} \mid L_{t-1} = k, RV^{t-1})\,\Pr(L_{t-1} = k \mid RV^{t-1}), \quad \text{(B11)}$$

where $p(\beta_{t-1}^{(k)} \mid L_{t-1} = k, RV^{t-1})$ is the likelihood function given by (12). For the sake of brevity, we denote $\pi_{t|s,k} = \Pr(L_t = k \mid RV^s)$ as the inclusion probability that M_k is included at time t given the information from beginning to period *s*. We define P as the transition matrix for the previous period for all predictors with elements $p_{ij} = \Pr(L_t = i \mid L_{t-1} = j), i, j = 1, \cdots, \mathrm{K}$. Then, we have $\pi_{t|t-1,k} = \sum_{l=1}^{K} \pi_{t-1|t-1,1} p_{kl}$. Raftery *et al*. (2010) estimate the inclusion probability by using the forgetting factor *α*. $\pi_{t|t-1,k}$ is estimated as follows:

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

$$\propto \left[\pi_{t-1|t-2,k}\, p_k(RV_{t-1} \mid RV^{t-2})\right]^{\alpha} = \prod_{i=1}^{t-1}\left[p_k(RV_{t-1} \mid RV^{t-i-1})\right]^{\alpha^i} \quad \text{(B12)}$$

And $\pi_{t|t,k}$ is updated with

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k}\, p_k(RV_t \mid RV^{t-1})}{\sum_{l=1}^{K} \pi_{t|t-1,l}\, p_l(RV_t \mid RV^{t-1})}, \quad \text{(B13)}$$

where $p_l(RV_t \mid RV^{t-1})$ is the prediction density for the l-th model conditional on the previous information defined in (15).