Journal of Management Science and Engineering 4 (2019) 55–73

Science Press

Contents lists available at ScienceDirect

Journal of Management Science and Engineering

journal homepage: www.keaipublishing.com/en/journals/journal-of-management-science-and-engineering/

Versatile HAR model for realized volatility: A least square model averaging perspective

Yue Qiu [a], Xinyu Zhang [b, d, *], Tian Xie [a], Shangwei Zhao [c]

[a] WISE and School of Economics, Xiamen University, Xiamen, Fujian, 361005, China
[b] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, China
[c] College of Science, Minzu University of China, Beijing, 100081, China
[d] Department of Statistics, Pennsylvania State University, University Park, Pennsylvania 16802, USA

A R T I C L E I N F O

Article history:
Available online 1 April 2019

Keywords:
Heterogeneous autoregression
Volatility forecasting
Forecasting combination
Model averaging
Asymptotic optimality

A B S T R A C T

A rapidly growing body of literature has documented improvements in forecasting financial return volatility measurement using various heterogeneous autoregression (HAR) type models. Most HAR-type models use a fixed lag index of $(1,5,22)$ to mirror the daily, weekly, and monthly components of the volatility process, but they ignore model specification uncertainty. In this paper, we propose applying the least squares model averaging approach to HAR-type models with signed realized semivariance to account for model uncertainty and to allow for a more flexible lag structure. We denote this approach as MARS and prove that the MARS estimator is asymptotically optimal in the sense of achieving the lowest possible mean squared forecast error. Selected by the data-driven model averaging method, the lag combination in the MARS method changes with various data series and different forecast horizons. Employing high frequency data from the NASDAQ 100 index and its 104 constituents, our empirical results demonstrate that acknowledging model uncertainty under the HAR framework and solving with the model averaging method can significantly improve the accuracy of financial return volatility forecasting.

1. Introduction

Modeling the volatility of financial assets is of great interest to many practitioners in finance. To capture the slowly decaying autocorrelation feature of this variance time series,[1] also known as long memory, various models have been suggested, such as the renowned fractionally integrated autoregressive moving average (ARFIMA) models in Andersen, Bollerslev, Diebold, and Labys (2001b) and the heterogeneous autoregressive (HAR) model proposed by Corsi (2009). The HAR model essentially claims that the conditional variance of the discretely sampled returns is a linear function of the lagged squared return over the identical return horizon in combination with the squared returns over longer and/or shorter return

* Corresponding author. Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, China.
E-mail addresses: yueqiu3@xmu.edu.cn (Y. Qiu), xinyu@amss.ac.cn (X. Zhang), xietian@xmu.edu.cn (T. Xie), swzhaomuc@163.com (S. Zhao).

[1] This phenomenon has been documented by Dacorogna, Műller, Nagler, Olsen, and Pictet (1993) and Andersen et al. (2001b) for the foreign exchange market and by Andersen, Bollerslev, Diebold, and Ebens (2001a) for stock market returns.

https://doi.org/10.1016/j.jmse.2019.03.003

horizons. Compared with the ARFIMA model, the HAR model has computational simplicity (e.g., OLS) and excellent out-of-sample performance.

Inspired by the success of HAR-type models, most work following Corsi (2009) has extended the HAR model in the direction of generalizing with jumps, leverage effects, and other nonlinear behaviors.[2] For example, Corsi, Pirino, and Renò (2010) used the C-Tz test to detect jumps and included them in the HAR model. Bollerslev, Litvinova, and Tauchen (2006) emphasized the importance of asymmetric leverage effects when modeling the volatility process. More complicated nonlinear effects, including structural breaks and regime-switches, were modeled by McAleer and Medeiros (2008) and Scharth and Medeiros (2009). Patton and Sheppard (2015) proposed modeling the equity price volatility using some measure named signed realized semivariances (HAR-RS), and they showed that future volatility is more closely related to the volatility of past negative returns.

The HAR model has an intuitive interpretation that agents with daily, weekly, and monthly trading frequencies perceive and respond to, altering the corresponding components of volatility.[3] Therefore, the lag structure of the HAR is usually fixed as $(1, 5, 22)$. However, the suitability of such a specification has been challenged. Craioveanu and Hillebrand (2012) utilized a parallel computing method to investigate all of the possible combinations of lags (chosen within a maximum lag of 250). Their results corroborated the conventional assumption of $(1, 5, 22)$.

Audrino and Knaus (2016) demonstrated that a model selection method, such as the least absolute shrinkage and selection operator (LASSO), cannot recover the implied AR lag terms within the pure HAR model.[4] However, they exhibit comparable forecasting performance. Audrino, Camponovo, and Roth (2015) used the adaptive LASSO inference theory to formally test the optimal lag structure. They suggested that the optimal significant lag structure is time-varying and subject to regime shifts. Audrino, Huang, and Okhrin (2016) employed the group LASSO to propose a flexible HAR model with changing time scales for HAR components. They discovered that the lag structure of the HAR model does exist, but the time scales for each component in the cascade differ from the original assumption $(1, 5, 22)$, particularly when the market environment is not stable.

We complement the above lines and contribute to the literature by studying model specification uncertainty under the HAR framework from a least squares model averaging perspective.[5] The potential model set is built based on the HAR-RS type models in Patton and Sheppard (2015) because of their sound out-of-sample performance. By taking a full permutation of all of the possible HAR lag terms to build the potential models, we can accommodate a dynamic lag structure in the HAR-RS type models, instead of restricting the lag index to $[1, 5, 22]$. Our method is termed MARS, short for model averaging HAR model with realized semivariance components. The MARS is data-driven, as the empirical weights on candidate models with different lag indices alter with the underlying data sets and forecast horizons. We also prove that the MARS estimator is asymptotically optimal in the sense of achieving the lowest possible mean squared forecast error.

In the empirical analysis, we consider applying our MARS method to model and forecast realized variance for a large set of assets—the NASDAQ 100 index and its 104 component stocks. In the out-of-sample exercise, we show that our MARS method exceeds the performance of the other nine popular volatility models, including the HAR-RS type models. Moreover, using the Giacomini-White test (Giacomini & White, 2006), we reveal that the MARS improvement is mostly significant at both the 10% and 5% levels. Especially for the longer forecast horizons of $h = 5, 10$, and 22, the MARS method completely dominates.

The remainder of the paper is organized as follows. Section 2 presents other comparison models. Section 3 proves the asymptotic optimality of the MARS method. Section 4.2 describes the data set and illustrates the empirical application to the NASDAQ 100 index and its component stocks. Section 5 concludes. A detailed proof for the asymptotic optimality of the MARS estimator can be found in Appendix A. Appendices B and C discuss the data set and the choice of exogenous variables. Appendix D provides additional empirical results.

2. A review of alternative models

In this section, we provide a review of the alternative models, which act either as a base model for our averaging method or as comparison models in our exercise. Following Andersen and Bollerslev (1998), the M-sample daily realized variance at day t (RV_t) can be calculated by summing the corresponding M equally spaced intra-daily squared returns $r_{t,j}$. Here, the subscript t indexes the day, and j indicates the time interval within day t,

[2] Corsi, Audrino, and Renò (2012) provided a comprehensive review of the development of HAR-type models and their possible extensions.

[3] Müller et al. (1993) referred to this idea as the Heterogeneous Market Hypothesis.

[4] The LASSO and its related methods have received considerable attention in recent decades since the seminal work of Tibshirani (1996). If the number of regressors exceeds the number of observations, LASSO can simultaneously shrink the number of regressors and obtain their coefficient estimates.

[5] Wang, Ma, Wei, and Wu (2016) proposed using a dynamic model averaging (DMA) approach that combines the forecasts of several HAR-type models to forecast the realized variance of the S&P 500 index. Different from our paper, they do not consider the implications of the HAR model specification uncertainty for out-of-sample forecasts. In addition, their exercise is limited to one-day ahead forecasts, while we study long-horizon forecasts such as 5-day, 10-day, and 22-day ahead forecasts. Their paper also differs from ours regarding the estimation method. The Bayesian method is adopted by their paper to estimate the time-varying coefficients, whereas we update the least square model averaging estimates and their relevant model weights. In a similar fashion, Liu and Maheu (2009) focused on Bayesian model averaging to exploit the benefits of including additional measures of volatility.

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

where $t = 1,2,\ldots,T$, $j = 1,2,\ldots,M$, and $r_{t,j}$ is the difference between log-prices $p_{t,j}$ ($r_{t,j} = p_{t,j} - p_{t,j-1}$).

Among the RV models, the HAR model introduced by Corsi (2009) has gained great popularity because of its estimation simplicity and outstanding out-of-sample performance. The standard HAR model in Corsi (2009) postulates that the h-step-ahead daily RV_{t+h} can be modeled by [6]

$$log\mathrm{RV}_{t+h} = \beta_0 + \beta_d log\mathrm{RV}_t^{(1)} + \beta_w log\mathrm{RV}_t^{(5)} + \beta_m log\mathrm{RV}_t^{(22)} + \varepsilon_{t+h}, \tag{2}$$

where the explanatory variables can take the general form of $log\mathrm{RV}_t^{(l)}$. $log\mathrm{RV}_t^{(l)}$ is defined by

$$log\mathrm{RV}_t^{(l)} \equiv l^{-1} \sum_{s=1}^{l} log\mathrm{RV}_{t-s} \tag{3}$$

as the l period averages of daily log RV, the βs are the coefficients, and $\{\varepsilon_t\}_t$ is a zero mean innovation process.

Since each $log\mathrm{RV}_t^{(l)}$ can be regarded as a volatility cascade, generated by the actions of distinct types of market participants trading at daily, weekly, or monthly frequencies (Müller et al., 1993), the lag structure in the HAR model is fixed at some lag index vector $\boldsymbol{l} = [1,5,22]$.

To control for the impacts of exogenous explanatory variables, we also consider the HAR model with exogenous variables (HARX), where a set of K-dimensional exogenous regressors $\boldsymbol{z}_t = [z_{1t},\ldots,z_{Kt}]$ is incorporated into Equation (2).

$$log\mathrm{RV}_{t+h} = \beta_0 + \beta_d log\mathrm{RV}_t^{(1)} + \beta_w log\mathrm{RV}_t^{(5)} + \beta_m log\mathrm{RV}_t^{(22)} + \boldsymbol{z}_t\beta_z + \varepsilon_{t+h}. \tag{4}$$

To further allow for asymmetric effects from exogenous variables on volatility dynamics, Fernandes, Medeiros, and Scharth (2014) suggested the asymmetric HARX (AHARX) model, which can be expressed by

$$log\mathrm{RV}_{t+h} = \beta_0 + \beta_d log\mathrm{RV}_t^{(1)} + \beta_w log\mathrm{RV}_t^{(5)} + \beta_m log\mathrm{RV}_t^{(22)} + \boldsymbol{z}_t^-\beta_z^- + \boldsymbol{z}_t^+\beta_z^+ + \varepsilon_{t+h}, \tag{5}$$

where $\boldsymbol{z}_t^- = \{\boldsymbol{z}_{1t}^-,\ldots,\boldsymbol{z}_{Kt}^-\}$ and $\boldsymbol{z}_t^+ = \{\boldsymbol{z}_{1t}^+,\ldots,\boldsymbol{z}_{Kt}^+\}$ for $k = 1,\ldots,K$, with

$$z_{kt}^- = \begin{cases} z_{kt}\mathbb{I}(z_{kt}<0) & \text{if a return} \\ z_{kt}\mathbb{I}(\Delta z_{kt}<0) & \text{if in levels} \end{cases} \qquad z_{kt}^+ = \begin{cases} z_{kt}\mathbb{I}(z_{kt}>0) & \text{if a return} \\ z_{kt}\mathbb{I}(\Delta z_{kt}>0) & \text{if in levels} \end{cases}$$

$\mathbb{I}(\cdot)$ is an indicator function that equals one if the argument of the function is satisfied.

To capture the information from signed high-frequency variation, Patton and Sheppard (2015) developed a series of HAR-RS type models. The first one, HAR-RS-I model, completely decomposes the $\mathrm{RV}^{(1)}$ in Equation (2) into two asymmetric semivariances, RS_t^+ and RS_t^-,

$$log\mathrm{RV}_{t+h} = \beta_0 + \beta_d^+ log\mathrm{RS}_t^+ + \beta_d^- log\mathrm{RS}_t^- + \beta_w log\mathrm{RV}_t^{(5)} + \beta_m log\mathrm{RV}_t^{(22)} + \varepsilon_{t+h}, \tag{6}$$

where $\mathrm{RS}_t^- = \sum_{j=1}^{M} r_{t,j}^2 \cdot \mathbb{I}(r_{t,j}<0)$ and $\mathrm{RS}_t^+ = \sum_{j=1}^{M} r_{t,j}^2 \cdot \mathbb{I}(r_{t,j}>0)$. To verify the actual effects of signed variations, they include an additional term capturing the leverage effect, $log\mathrm{RV}_t^{(1)} \cdot \mathbb{I}(r_t<0)$. The second model in Equation (7) is denoted as HAR-RS-II,

$$\begin{aligned} log\mathrm{RV}_{t+h} = \beta_0 + \beta_1 log\mathrm{RV}_t^{(1)} \cdot \mathbb{I}(r_t<0) + \beta_d^+ log\mathrm{RS}_t^+ + \beta_d^- log\mathrm{RS}_t^- + \beta_w log\mathrm{RV}_t^{(5)} \\ + \beta_m log\mathrm{RV}_t^{(22)} + \varepsilon_{t+h}. \end{aligned} \tag{7}$$

The third and fourth models in Patton and Sheppard (2015), denoted as HAR-SJ-I (Equation (8)) and HAR-SJ-II (Equation (9)), respectively, examine the role that decomposing realized variances into signed jump variations and bipower variation (BPV) can play in forecasting volatility:

$$log\mathrm{RV}_{t+h} = \beta_0 + \beta_d^j \%\mathrm{SJ}_t + \beta_d^{bpv} log\mathrm{BPV}_t + \beta_w log\mathrm{RV}_t^{(5)} + \beta_m log\mathrm{RV}_t^{(22)} + \varepsilon_{t+h}, \tag{8}$$

[6] Using the log to transform the realized variance is standard in the literature, motivated by avoiding imposing positive constraints and worrying about the heteroskedasticity of residuals relative to the level of the process, as mentioned by Patton and Sheppard (2015). An alternative is to implement weighted least squares (WLS) on RV, which is not well suited to our purpose of using the least squares model averaging method.

$$
\begin{aligned}
log\mathrm{RV}_{t+h} = \beta_0 + \beta_d^{j-}\%\mathrm{SJ}_t^- + \beta_d^{j+}\%\mathrm{SJ}_t^+ + \beta_d^{bpv} log\mathrm{BPV}_t + \beta_w log\mathrm{RV}_t^{(5)} \\
+\beta_m log\mathrm{RV}_t^{(22)} + \varepsilon_{t+h},
\end{aligned} \tag{9}
$$

where signed jump variation SJ_t is derived from $\mathrm{SJ}_t = \mathrm{RS}_t^+ - \mathrm{RS}_t^-$. The quadratic variation is decomposed into signed jump variation, $\%\mathrm{SJ}_t = log\left(1 + \frac{SJ_t}{RV_t}\right)$, and its continuous component using bipower variation, $log\mathrm{BPV}_t \equiv log(2/\pi)^{-1}\sum_{j=2}^{M}|r_{t,j-1}||r_{t,j}|$. $\%\mathrm{SJ}_t^-$ and $\%\mathrm{SJ}_t^+$ further decompose $\%\mathrm{SJ}_t$ using an indicator variable for the sign of the difference, where $\%\mathrm{SJ}_t^- = \%\mathrm{SJ}_t\mathbb{I}(\mathrm{RS}_t^+ - \mathrm{RS}_t^- < 0)$. The HAR-SJ-II model distinguishes the effect of a positive jump variation from that of a negative jump variation.

The setup of $\boldsymbol{l} = [1, 5, 22]$ has been challenged recently in the model selection literature. Audrino and Knaus (2016) proved that LASSO would recover the same active lag terms in the AR framework as those in the HAR model asymptotically only if the HAR model was the true data generating process (DGP).[7] Audrino et al. (2016) put forward the flexible HAR (FHAR) that selects active lag terms from a full lag index vector $\boldsymbol{l} = [1, 2, \ldots, L]$ using the adaptive LASSO. [8] Also, Audrino et al. (2016) find in their empirical exercise that the FHAR is data-driven and comparatively resilient to the unstable market environment.

For the estimation of FHAR, we consider the following equation with a full lag index vector $\boldsymbol{l} = [1, 2, \ldots, L]$

$$
log\mathrm{RV}_{t+h} = \beta_0 + \beta_1 log\mathrm{RV}_t^{(1)} + \beta_2 log\mathrm{RV}_t^{(2)} + \ldots + \beta_L log\mathrm{RV}_t^{(L)} + \varepsilon_{t+h} \tag{10}
$$

subject to the constraint $\sum_{l=1}^{L}|\beta_l| < c$ for some constant c. The coefficient matrix is $\boldsymbol{\beta} = [\beta_0, \beta_1, \ldots, \beta_L]^\top$. Following Audrino et al. (2016), the adaptive LASSO estimator shrinks the OLS estimate $\widehat{\beta}$ to $\widehat{\beta}^{\mathrm{AL}}$ by solving

$$
\widehat{\beta}^{\mathrm{AL}} = \underset{\beta}{\mathrm{argmin}} \sum_{t=L}^{T} \left\{ log\mathrm{RV}_{t+h} - \left[1, log\mathrm{RV}_t^{(1)}, \ldots, log\mathrm{RV}_t^{(L)}\right]\boldsymbol{\beta} \right\}^2 + \lambda \sum_{l=1}^{L} \lambda_l \left|\beta_l\right|, \tag{11}
$$

where L is the maximum lag order, which is chosen as 50.[9] λ is the tuning parameter that controls the strictness of the penalty term.[10] In practice, researchers either assign λ a specific value or use s-fold cross-validation to determine the optimal λ. Here we use λ to minimize a five-fold cross-validation. The weights λ_l for each coefficient are calibrated as the inverse of the absolute value of the corresponding preliminary ridge regression estimator, as suggested by Audrino et al. (2016).

In practice, RV_{t+h} is unobservable at time t. Hence, both the dependent and explanatory variables of the models in Section 2 must take h-period of lags.

3. Model averaging the HAR-RS type models

This section establishes the motivation for model averaging the HAR-RS type models that follow. We then introduce the estimation procedure via the model averaging approach. Finally, the theoretical foundation of the asymptotic optimality of the proposed approach is briefly discussed.

3.1. *The model averaging estimator*

There has been some pioneering work, from a model selection perspective, on testing the validity of the lag structure in the conventional HAR model; see, e.g., Audrino et al. (2015), Audrino and Knaus (2016), Audrino et al. (2016), and others. While the lag terms in the HAR model survive the tests based on LASSO and the adaptive LASSO (Audrino et al., 2015; Audrino & Knaus, 2016), there is strong evidence in Audrino et al. (2016) that casts some doubt on the fixed choice of aggregation frequencies in the HAR model. In particular, Audrino et al. (2016) found that a fixed lag structure, i.e., $(1, 5, 22)$, is not statistically sustained by the group LASSO estimates for individual stocks under unstable market environments, such as the 2007–2009 crisis. They addressed the above issue with a proposed flexible HAR model, built completely from a model-selection point of view.

To answer the above question from a different angle, we consider the forecast implications of a flexible lag structure generated by the model averaging method. The potential model set is built upon the HAR-RS type models in Patton and

[7] Note that the theoretical results of Audrino and Knaus (2016) are only applicable to the HAR model in the form of (2). Extending the results of Audrino and Knaus (2016) to HAR models with additional components such as semivariances or jumps is beyond the scope of this paper and left for further research.

[8] The LASSO method is aimed at selecting a subset of important covariates by shrinking the coefficients of redundant ones towards zero. Model averaging, on the other hand, computes a weighted average across a group of candidate models without explicitly targeting the covariate importance. Although there are previous studies (for example, Burnham & Anderson, 2002) suggesting an assessment of the relative importance of the covariates by ranking their respective cumulative model averaging weights, such a procedure still needs a theoretical verification. However intriguing, we leave this for future research.

[9] The setup of parameter L draws from Audrino et al. (2016). Our conclusions do not change with alternative choices of L.

[10] When $\lambda = 0$, $\widehat{\beta}^{\mathrm{AL}}$ is equivalent to the OLS estimator. Increasing λ will increase the penalty on explanatory variables with nonzero coefficients.

Table 1
The specification of regressors for MARS.

Regressor Set-up		Reference
(i)	$[1, logRS^{-}_{t-h}, logRS^{+}_{t-h}, logRV^{(l_2)}_{t-h}, \ldots, logRV^{(l_p)}_{t-h}]$	HAR-RS-I
(ii)	$[1, logRV^{(1)}_{t-h}\mathbb{I}_{r_{t-h}<0}, logRS^{-}_{t-h}, logRS^{+}_{t-h}, logRV^{(l_2)}_{t-h}, \ldots, logRV^{(l_p)}_{t-h}]$	HAR-RS-II
(iii)	$[1, \%SJ_{t-h}, logBPV_{t-h}, logRV^{(l_2)}_{t-h}, \ldots, logRV^{(l_p)}_{t-h}]$	HAR-SJ-I
(iv)	$[1, \%SJ^{-}_{t-h}, \%SJ^{+}_{t-h}, logBPV_{t-h}, logRV^{(l_2)}_{t-h}, \ldots, logRV^{(l_p)}_{t-h}]$	HAR-SJ-II

Sheppard (2015), because the HAR-RS type models are shown to exhibit a significantly better out-of-sample forecast performance for a wide class of assets, compared with other HAR models.

Suppose that the dependent variable is $\boldsymbol{y} = [logRV_1, \ldots, logRV_T]^\top$ and the explanatory variable is $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T]^\top$,[11] where the specification of $\boldsymbol{x}_t$ is presented in the following Table 1.

It is obvious that the $\boldsymbol{x}_t$ variable is formed using one of the four HAR-RS type models in Patton and Sheppard (2015). Here, we do not restrict the lag index vector $\boldsymbol{l} = [l_2, \ldots, l_p]$ to $[5, 22]$. Instead, we acknowledge the specification uncertainty in $\boldsymbol{l}$ and consider a group of M candidate models to approximate the true data generating process. To achieve this, a set of lag index vectors is first produced by taking a full permutation of all the l_k s from $[l_2, \ldots, l_p] = [2, \ldots, 22]$. The maximum lag order l_p is chosen as 22.[12] Then, a group of M candidate models is constructed based on each $\boldsymbol{l}$ in the generated lag index set.[13] After model averaging, there will be distinct model weights assigned to candidate models based on different lag index vectors. Moreover, as the underlying data sets vary, this can alter the relevant model weights, making our method dynamic and data-driven.

Note that the model averaging estimator with screened candidate models is implemented in this paper, since keeping the total number of candidate models small or slowing its convergence to infinity is a necessary condition to maintain the asymptotic optimality of the model averaging estimators. However, in the context of the HAR model with the maximum lag order of l_p, we could end up with 2^{l_p} candidate models, and the number of potential models grows exponentially with l_p (Audrino & Knaus, 2016; Craioveanu & Hillebrand, 2012). To circumvent this issue, we can apply the model screening method, for example, the "top m" approach by Yuan and Yang (2005) and Zhang, Lu, and Zou (2013a) in which m is the number of candidate models after screening or the heteroscedasticity-robust model screening approach by Xie (2017), which can shrink the number of potential models before model averaging to an appropriate degree. For brevity, we call our model estimator, averaging the HAR-RS type models, the MARS estimator.

We define the true model as

$$\boldsymbol{y} = \boldsymbol{\mu} + \varepsilon,$$

where $\boldsymbol{y} = [y_1, \ldots, y_T]^\top$, $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_T]^\top$, and $\varepsilon = [\varepsilon_1, \ldots, \varepsilon_T]^\top$. μ_t can be considered as the conditional mean in period t, $\mu_t = \mathbb{E}(y_t | y_{t-h}, y_{t-h-1}, \ldots)$, and the error term ε_t has zero conditional mean $\mathbb{E}(\varepsilon_t | y_{t-h}, y_{t-h-1}, \ldots) = 0$ and i.i.d. variance σ^2. Let the m^{th} candidate model be

$$\boldsymbol{y} = \boldsymbol{X}^m \beta^m + \varepsilon^m,$$

where $\boldsymbol{X}^m$ are subsets of columns of $\boldsymbol{X}$. With $\boldsymbol{X}^m$ at hand, β^m can be estimated by $\widehat{\beta}^m = (\boldsymbol{X}^{m\top}\boldsymbol{X}^m)^{-1}\boldsymbol{X}^{m\top}\boldsymbol{y}$; therefore, $\boldsymbol{\mu}$ is approximated by $\widehat{\boldsymbol{\mu}}^m = \boldsymbol{X}^m\widehat{\beta}^m$. Following Hansen (2008), the optimal mean square h-period ahead forecast is the conditional mean μ_{T+h}. Therefore, the least-squares forecast of y_{T+h} from the m^{th} approximation model equals $\widehat{y}^m_{T+h} = \widehat{\mu}^m_{T+h} = \boldsymbol{x}^{m\top}_{T+h}\widehat{\beta}^m$, where $\boldsymbol{x}^m_{T+h}$ is assumed to follow $\boldsymbol{x}^m_{T+h} = \boldsymbol{x}^m_T$ and becomes observable in period t.

We obtain the forecasts of y_{T+h} from all approximation models and define the vector of forecasts $\widehat{\boldsymbol{y}}_{T+h}$

$$\widehat{\boldsymbol{y}}_{T+h} \equiv \left[\widehat{y}^1_{T+h}, \widehat{y}^2_{T+h}, \cdots, \widehat{y}^M_{T+h}\right]^\top. \tag{12}$$

The model averaging forecast is simply the weighted average of $\widehat{\boldsymbol{y}}_{T+h}$ such that

$$\widehat{y}_{T+h}(\boldsymbol{w}) \equiv \boldsymbol{w}^\top \widehat{\boldsymbol{y}}_{T+h} = \sum_{m=1}^{M} w^m \widehat{y}^m_{T+h},$$

where $\boldsymbol{w} = [w^1, \ldots, w^M]^\top$ is a weight vector in the unit simplex in $\mathbb{R}^M$

[11] Although all of the elements in $\boldsymbol{x}_t$ are h-period lags from period t, we follow the conventional notation in time series and denote $\boldsymbol{x}_t$ as the explanatory variable corresponding to the period t dependent variable.

[12] Our results are not sensitive to the arbitrary choice, as mentioned in Section 4.2.1 and Audrino et al. (2016).

[13] In this way, we expect to see that the candidate models will have full-fledged combinations of $logRV^{(l_k)}_{t-h}$.

$$\mathscr{H} \equiv \left\{ \boldsymbol{w} \in [0,1]^M : \sum_{m=1}^{M} w^m = 1 \right\}.$$

The performance of model averaging forecasts depends crucially on the weight vector $\boldsymbol{w}$. In this paper, we propose using the weight vector $\boldsymbol{w}$ selected by the MARS estimator. We define $\widehat{\boldsymbol{\mu}}^m = \boldsymbol{X}^m \widehat{\beta}^m$ as the vector of the estimated conditional mean by Model m. Let $\boldsymbol{X}$ be the regressor matrix of the full model with all regressors, so a project matrix $\boldsymbol{\Pi}^m$ exists such that $\boldsymbol{X}^m = \boldsymbol{X}\boldsymbol{\Pi}^m$. We define the coefficient estimates $\widehat{\beta}(\boldsymbol{w})$ by

$$\widehat{\beta}(\boldsymbol{w}) = \sum_{m=1}^{M} w^m \boldsymbol{\Pi}^{m\top} \widehat{\beta}^m.$$

The model averaging estimator of the conditional mean $\widehat{\boldsymbol{\mu}}(\boldsymbol{w})$ is therefore defined by

$$\widehat{\boldsymbol{\mu}}(\boldsymbol{w}) \equiv \boldsymbol{X}\widehat{\beta}(\boldsymbol{w}). \tag{13}$$

Like most model selection and model averaging criteria, the MARS criterion balances the fit and complexity of a model:

$$\mathrm{MARS}(\boldsymbol{w}) = (\boldsymbol{y} - \widehat{\boldsymbol{\mu}}(\boldsymbol{w}))^\top (\boldsymbol{y} - \widehat{\boldsymbol{\mu}}(\boldsymbol{w})) \left(\frac{T + k(\boldsymbol{w})}{T - k(\boldsymbol{w})} \right), \tag{14}$$

where $k(\boldsymbol{w}) \equiv \sum_{m=1}^{M} w^m k^m$ is the effective number of parameters and k^m is the number of regressors in model specification $\boldsymbol{X}^m$. Assume that k^m is unrelated to T. Criterion (14) can be used to calculate the empirical weight vector $\widehat{\boldsymbol{w}}$, in which

$$\widehat{\boldsymbol{w}} = \underset{\boldsymbol{w} \in \mathscr{H}}{\arg\min}\, \mathrm{MARS}(\boldsymbol{w}).$$

Then, we calculate the model averaging forecast of y_{T+h} following

$$\widehat{y}_{T+h}(\widehat{\boldsymbol{w}}) = \widehat{\boldsymbol{w}}^\top \widehat{\boldsymbol{y}}_{T+h},$$

where $\widehat{\boldsymbol{y}}_{T+h}$ is the vector of conditional mean estimates from M approximation models. Note that our MARS estimator is similar to the PMA estimator of Xie (2015), whereas the original PMA estimator does not assume a dynamic model structure. Hence, the MARS estimator can be considered an extension of the PMA estimator under the HAR-RS framework.

3.2. Asymptotic optimality of MARS estimator

In this section, we prove the asymptotic optimality of MARS in the sense of minimizing the squared estimation loss $L_T(\boldsymbol{w}) \equiv \|\widehat{\boldsymbol{\mu}}(\boldsymbol{w}) - \boldsymbol{\mu}\|^2$. The following regularity conditions are required to complete the proof:

(C.1) For any $m \in \{1, \ldots, M\}$, there exists $\tilde{\beta}^m$ such that $\widehat{\beta}^m - \tilde{\beta}^m = O_p(T^{-1/2})$ uniformly.

(C.2) Let $\tilde{\beta}(\boldsymbol{w}) = \sum_{m=1}^{M} w^m \boldsymbol{\Pi}^{m\top} \tilde{\beta}^m$, $\tilde{L}_T(\boldsymbol{w}) = \left\| \boldsymbol{X}\tilde{\beta}(\boldsymbol{w}) - \boldsymbol{\mu} \right\|^2$, and $\xi_T = \inf_{w \in \mathscr{H}} \tilde{L}_T(\boldsymbol{w})$. It is assumed that $\xi_T^{-1} T^{1/2} M^2 = o_p(1)$ and $\max_{m \in \{1, \ldots, M\}} k_m \xi_T^{-1} = o_p(1)$.

(C.3) $\boldsymbol{X}^\top \varepsilon = O_p(T^{1/2})$, $\boldsymbol{X}^\top \boldsymbol{X}/T = O_p(1)$, and $\boldsymbol{\mu}^2/T = O_p(1)$.

Condition (C.1) is a high-level condition and is imposed on the convergence of the estimators under misspecification. White (1982) showed that, for misspecified models, the parameter estimators converge to their limits with root T rate under some reasonable assumptions. Condition (C.2) is implied by condition (11) of Ando and Li (2014) and requires that the minimum limiting squared loss expands at a rate more rapid than $T^{1/2}M^2$ and $\max_{m \in \{1, \ldots, M\}} k_m$, essentially making this condition a restriction on the degrees of misspefication of candidate models. Condition (C.2) also needs $M = o(T^{-1/4}\xi_T^{1/2})$, which is a restriction of M, but allows M to diverge with T when $\xi_T \to \infty$ with a proper rate. The first two parts of Condition (C.3) are high-level and allow for the application of cross-sectional and time-series data. The third part of Condition (C.3), regarding the average of μ_i^2, is a similar condition adopted by Zhang, Wan, and Zou (2013b) and Xie (2015).

Given the above Conditions (C.1) to (C.3), we can demonstrate the asymptotic optimality of $\widehat{\boldsymbol{w}}$ in Theorem 1.

Table 2
Summary statistics of the RV series for the NASDAQ 100 index.

Sample Statistics	First Half	Second Half	Full Sample
Mean	1.5485	1.5239	1.5388
Median	0.7606	0.7074	0.7247
Maximum	58.7534	53.2069	58.7534
Minimum	0.0531	0.0491	0.0491
Standard Deviation	2.4862	2.5605	2.5088
Skewness	6.6979	6.3115	6.3498
Kurtosis	91.5761	74.1242	78.3111
Jarque-Bera	0.0010	0.0010	0.0010
ADF	0.0010	0.0010	0.0010

The second column of the table represents the subsample period ranging from January 7, 1998 to July 6, 2007, while the third column represents the subsample period ranging from July 9, 2007 to March 24, 2017. The fourth column contains statistics for the entire time period from January 7, 1998 to March 24, 2017, for a total of 1,810,756 observations. Table 2 also reports the p-values of the Jarque-Bera and ADF tests at 0.001.

Theorem 1. *Assuming Conditions (C.1) to (C.3) hold, then we have*

$$\frac{L_T(\widehat{\boldsymbol{w}})}{\inf_{\boldsymbol{w}\in\mathscr{H}} L_T(\boldsymbol{w})} \to 1, \tag{15}$$

when $T\to\infty$.

The complete proof of the above theorem is given in Appendix A. Equation (15) implies that, using $\widehat{\boldsymbol{w}}$, the MARS method has a squared loss asymptotically identical to that of the infeasible best possible model average estimator. In other words, the MARS method is optimal.

4. Empirical application

4.1. Data description

The data set used for the empirical analysis consists of high-frequency transaction prices from the NASDAQ 100 index between January 7, 1998 and March 24, 2017.[14] We also study the 104 constituents of the NASDAQ 100 Index for comparison. The constituents cover industry groups ranging from computer hardware and software, telecommunications, retail/wholesale trade, to biotechnology. A more detailed description of these stocks is given in Appendix B, including their ticker symbols, names, and Global Industry Classification (GIC) sectors. Although the start date of each individual stock varies,[15] they all possess the same end date of March 24, 2017. The whole data set and its relevant information are taken from the New York Stock Exchange's TAQ database and Datastream. The data on the NASDAQ 100 index and individual stock prices are both obtained at 1-min increments. The intraday prices are then used to calculate daily realized variance measures based on the two-time scales estimator by Zhang, Mykland, and Aït-Sahalia (2005). The two selected scales are 2 and 20 ticks; however, the results are not sensitive to the grid size.

We describe summary statistics of the realized variance series for the NASDAQ 100 index in Table 2. Our intraday data span from January 7, 1998 to March 24, 2017 with 1-min increments, totaling 1,810,756 observations. To compare the changes that occur over the subsample periods, the full sample is divided into two parts. In each column, Table 2 documents the results of the sample mean, median, minimum, maximum, standard deviation, skewness, and kurtosis for the realized variance series over the two subsample periods and the full sample period. As can be seen from Table 2, the first period shows slightly larger values over all of the statistics, with the only exception being standard deviation. It should be noted that the first is the period corresponding to the aftermath of the burst of the tech bubble (the fourth quarter of 2002) and the terrorist attacks on 9/11. It is not surprising that the first period seems to be more volatile, if we consider the proportion of information technology companies in the NASDAQ 100 index (i.e., 46 of 104 companies).

Table 2 also reports the *p*-values[16] of the Jarque-Bera test for normality and those of the augmented Dickey-Fuller (ADF) test for unit root. The null hypotheses of a normal distribution and a unit root are strongly rejected by both the Jarque-Bera and ADF tests in the two subsamples as well as in the full sample.

In Table 3, we also report summary statistics for the realized variance series of the top 10 value weighted components in the NASDAQ 100 index. Their respective weights in percentage terms are listed in the last row. The *p*-values of the Jarque-Bera and ADF tests for the ten stocks are recorded respectively to test their normality and stationarity. The findings show that we

[14] Note that we start our sample after US equities were traded with a spread of less than one-eighth of a dollar. Aït-Sahalia and Yu (2009) argued that this condition could change the market noise level and affects the estimator of realized variance, although the effect was not significant.

[15] Their start dates vary from January 2, 1990 to December 16, 2010.

[16] In our exercises, we set the lower bound of the *p*-values of the Jarque-Bera and ADF tests at 0.001. Values less than 0.001 are truncated at 0.001.

Table 3
Summary statistics of the RV series for selective stocks in the NASDAQ 100 index.

Sample Statistics	AAPL	AMZN	MSFT	FB	GOOG	GOOGL	INTC	CSCO	CMCSA	NVDA
Mean	5.5155	4.1996	2.1830	3.2346	1.3842	2.5875	3.0214	2.8515	3.4724	7.1237
Median	2.3075	2.7976	1.4353	2.0220	0.9631	1.4460	2.0897	1.9765	2.2267	5.0260
Maximum	189.2285	134.1444	84.4586	65.7628	36.9286	101.7953	107.7959	126.0908	160.1925	165.9007
Minimum	0.1571	0.1472	0.1753	0.1426	0.1536	0.1958	0.2812	0.0497	0.2510	0.5610
Standard Deviation	9.1409	6.2147	3.1466	4.3227	1.7785	4.3324	3.8780	4.1806	5.1658	7.8537
Skewness	5.7862	8.5128	9.9314	6.1704	10.2482	9.2924	9.3777	12.1930	11.1117	5.6521
Kurtosis	71.6176	114.6391	175.6459	67.1584	183.7117	148.9310	177.1774	262.3127	256.4778	68.9745
Jarque-Bera Test	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
ADF Test	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
Weights in NASDAQ 100	11.3240	9.7810	9.1320	5.1310	4.8910	4.1660	3.0650	2.7220	2.0410	1.9230

The table reports the mean, median, minimum, maximum, standard deviation, skewness, and kurtosis for the realized variance series of the top 10 value weighted components in the NASDAQ 100 index. Their respective weights in percentage terms are listed in the last row. The p-values of the Jarque-Bera and ADF tests for the ten stocks are recorded respectively, in order to test their normality and stationarity.

can still reject the null hypotheses of a normal distribution and unit root, very similar to what is uncovered for the NASDAQ 100 index.

4.2. Rolling window analysis

In this section, we conduct several empirical exercises to compare the out-of-sample forecasting performance of the MARS method with the usually used autoregressive model (ARX) and a battery of HAR-type models introduced in Section 2. The related models are listed as follows:

(i) ARX model: the simple autoregression model AR(22) with exogenous variables $\boldsymbol{z}_t$;
(ii) HAR model: defined in (2);
(iii) HARX model: defined in (4);
(iv) AHARX model: defined in (5);
(v) HAR-RS-I model: defined in (6);
(vi) HAR-RS-II model: defined in (7);
(vii) HAR-SJ-I model: defined in (8);
(vii) HAR-SJ-II model: defined in (9);
(ix) Flexible HAR model (FHAR): proposed by Audrino et al. (2016); and
(x) MARS method: introduced in Section 3.

Note that the ARX, HARX, and AHARX models all contain exogenous variables $\boldsymbol{z}_t$. Inspired by Fernandes et al. (2014), we include the following exogenous variables $\boldsymbol{z}_t$ both contemporaneously and with one lag: (i) the κ-day continuously compounded return on the NASDAQ 100 index for $\kappa = 1, 5, 10, 22, 66$; (ii) the first difference of the logarithm of the volume of the NASDAQ 100 index (NASDAQ 100 vol change); (iii) the j-day continuously compounded return on the one-month crude oil futures contract for $j = 1, 5, 10, 22, 66$ (oil j-day return); (iv) the first difference of the logarithm of the trade-weighted average of the foreign exchange value of the US dollar index against the Australian dollar, Canadian dollar, Swiss franc, euro, British pound sterling, Japanese yen, and Swedish krona (USD change); (v) the excess yield of the Moody's seasoned Baa corporate bond over the Moody's seasoned Aaa corporate bond (credit spread); (vi) the difference between the 10-Year and 3-month Treasury constant maturity rates (term spread); and (vii) the difference between the effective and target federal fund rates (FF deviation). Descriptive statistics for $\boldsymbol{z}_t$ are available in Appendix C.

As argued by Fernandes et al. (2014), multiperiod returns on the NASDAQ 100 index in $\boldsymbol{z}_t$ are included to account for the possible leverage and volatility feedback effects (Bollerslev & Zhou, 2006). Moreover, there is plentiful evidence emphasizing the volatility-volume relation (Becker, Clements, & White, 2007; Chan & Fong, 2000; Lamoureux & Lastrapes, 1990), motivating us to add the NASDAQ 100 vol change to $\boldsymbol{z}_t$. While both oil prices and term spread are concerned with different dimensions of the overall market conditions, USD change and FF deviation are both linked to US macroeconomic conditions.

As we mentioned in Section 3, for model averaging, a full permutation of all the potential variables may result in an overwhelming number of candidate models. In our exercises, we employ a modified ARMS approach to somewhat control the total number of candidate models. Following Yuan and Yang (2005) and Zhang et al. (2013a), the "top m" approach starts with a preliminary candidate model set that contains all the possible candidate models. We then sort the models by their values of Mallows' C_p from the lowest to the highest and (arbitrarily) select the top L models as the final candidate model set.[17] In our

[17] As discussed in Lehrer and Xie (2017), for least squares model averaging, it is quite common in practice for a handful of models to accumulate a large proportion of model weights and the rest are assigned near zero weights. In this case, models with near zero weights can be discarded without harming the results.

exercises, we eventually set $L = 20$ and discover that the sets of model candidates vary with the underlying datasets and different forecasting horizons. Other alternatives of L larger than 20 have been tested, which produce fairly similar forecast results. However, we do find out that choosing less than 20 for L yields worse outcomes.

For all of the exercises, we conduct a rolling window out-of-sample exercise. The window length is set at 1000. We also use other values for the window length, and the results are similar. The maximum lag order is chosen as 22.[18] Each of the above 10 models is applied to the data set, and a series of h periods ahead forecasts are obtained.[19] We consider both short-horizon and long-horizon forecasts with $h = 1, 5, 10$, and 22. By comparing the forecasts between models, we specifically estimate the below four statistics as the performance measure: (i) the mean squared forecast error (MSFE); (ii) the standard deviation of forecast error (SDFE); (iii) the mean absolute forecast error (MAFE); and (iv) the Mincer-Zarnowitz pseudo-R^2 for each model at each forecast horizon.

4.2.1. Forecasting the RV of the NASDAQ 100 index

Here, the MARS method is used to forecast the RV of the NASDAQ 100 index. We set the potential regressors in our MARS method according to Table 1, which further indicates that the set of candidate models for model averaging is based on one of the HAR-RS specifications and a full permutation of its lags from $log\text{RV}_{t-h}^{(l_2)}$ till $log\text{RV}_{t-h}^{(l_p)}$ ($\boldsymbol{l} = [1, 2, \ldots, 22]$). We end up choosing the HAR-RS-II as the base model for constructing the candidate model set. The possibility of using other HAR-RS type models as the base model is also explored in Appendix D. The results there clearly confirm that the HAR-RS-II outperforms all of the other HAR-RS type models for predicting the NASDAQ 100 index at all four horizons.[20] Nevertheless, it should be noted that, in terms of the MAFE and R^2, the MARS method based on other HAR-RS specifications still outperforms the comparison models without model averaging for three of the four horizons ($h = 5$, $h = 10$ and $h = 22$).

Table 4 exhibits some descriptive results of the out-of-sample evaluation for forecasts 1, 5, 10, and 22 days ahead. In particular, we report the MSFE, SDFE, and MAFE, as well as pseudo R^2 from the rolling-window exercise for all 10 models previously discussed. The first column of Table 4 lists the model specifications, including the MARS method specified above, apart from other HAR-type models and an AR model with exogenous variables. Columns 2–5 cover the results for several forecast criteria: the MSFE, SDFE, MAFE, and pseudo R^2, respectively. Panels A to D demonstrate various forecast horizons ranging from 1 to 22 days. Bold numbers indicate the best performing model by each criterion at each forecast horizon.

We note that, when the forecast horizon expands, the forecast errors of all models also increase from Panels A to D. To predict the NASDAQ 100 index volatility, it seems that imposing the HAR framework indeed helps. Except at $h = 1$ and $h = 5$, the HAR model is mildly exceeded by the ARX model by checking the MSFE, SDFE, and R^2. For the remaining forecast horizons, the ARX model performs poorly in general. The AHARX and HARX models have a mixed performance. At $h = 1$, both of them perform better than the HAR model. However, the pure HAR model has better forecast accuracy at longer forecast horizons ($h = 5, 10, 22$), which might imply that the extraneous variables $\boldsymbol{z}_t$ are, in this case, more useful for short-term forecasts and do not contribute much to long-term forecasts.

Relative to baseline HAR specifications, the semivariance based alternatives have superior performance, which is particularly true for $h = 5, 10, 22$ and which agrees with the conclusions in Patton and Sheppard (2015). Among the four HAR-RS type models, the HAR-RS-II model is in a dominant position for predicting the NASDAQ 100 volatility. However, the HAR-RS-II model is completely overcome by its model averaging counterpart (the MARS method) for all forecast horizons. The MARS alternative generates gains in out-of-sample R^2 of between 4.32% ($h = 1$) and 29.6% ($h = 10$), as well as of the absolute forecast errors of between 1.85% (h = 1) and 17.5% (h = 22).

Although both the Flexible HAR and our MARS method acknowledge model specification uncertainty, the model averaging based MARS again exceeds the performance of model selection based Flexible HAR at each horizon and by each criterion. We note that the Flexible HAR model yields similar results to those of the pure HAR model, consistent with the findings by Audrino et al. (2016). To summarize, our MARS method shows the best performance among all of the models considered at each forecast horizon. The above findings confirm that, for HAR type models, acknowledging the lag structure uncertainty with the model averaging method can remarkably improve the forecasting accuracy.

To examine whether the out-performance is statistically significant, we perform the modified Giacomini-White (GW) test (Giacomini & White, 2006) for the null hypothesis that the *row method* performs equally well as the *column method* in terms of the mean absolute forecast error. The corresponding p values are reported in Table 5, where each panel stands for a specific forecast horizon varying from 1 to 22 days ahead. A large p value signifies that the performance of the corresponding row and column methods are statistically indifferent, and a small p value indicates that the row method performs significantly different from the column method.

The last row of each panel shows that the p-values of the GW test statistic are less than the 5% level across all of the forecast horizons, except for three cases at $h = 1$. This outcome indicates that the out-performance of the MARS method is statistically significant for the NASDAQ 100 index. In the situation of no rejections at the 5% level, two of the three p-values of the GW test remain smaller than the 10% level. The only exception is the HAR-RS-II model at $h = 1$, for which the MARS method

[18] We also tried the lag index $\boldsymbol{l} = [1, 2, \ldots, 66]$, and the results are quite similar to what we present here.

[19] Note that, to compute h-day ahead forecasts, we employ a direct forecast approach in which we replace $log\text{RV}_{t+1}$ with $log\text{RV}_{t+h}$ in the above models. This approach permits us to produce multi-step ahead forecasts without imposing any assumption about future realizations on the explanatory variables.

[20] This point contrasts with the results in Patton and Sheppard (2015), in which the HAR-SJ-II model generally performs the best for the SPDR series.

Table 4
Out-of-sample forecast comparison of models for the RV of the NASDAQ 100 index.

Model	MSFE	SDFE	MAFE	Pseudo R^2
Panel A:h = 1				
ARX	2.5022	1.5818	0.6370	0.5722
HAR	2.5414	1.5942	0.5332	0.5651
HARX	2.3980	1.5486	0.5994	0.5901
AHARX	2.2771	1.5090	0.5867	0.6107
HAR-RS-I	2.4651	1.5701	0.5182	0.5782
HAR-RS-II	2.0305	1.4250	0.5068	0.6527
HAR-SJ-I	2.4225	1.5564	0.5172	0.5857
HAR-SJ-II	2.4271	1.5579	0.5238	0.5849
FHAR	2.5271	1.5897	0.5381	0.5671
MARS	**1.8640**	**1.3653**	**0.4973**	**0.6809**
Panel B:h = 5				
ARX	3.9772	1.9943	0.8086	0.2202
HAR	4.1185	2.0294	0.7228	0.2938
HARX	4.6900	2.1656	0.7828	0.1980
AHARX	4.6357	2.1531	0.7935	0.2070
HAR-RS-I	3.8873	1.9716	0.7112	0.3336
HAR-RS-II	3.8201	1.9545	0.7069	0.3454
HAR-SJ-I	3.8957	1.9737	0.7128	0.3322
HAR-SJ-II	3.8305	1.9572	0.7066	0.3434
FHAR	4.0721	2.0179	0.7281	0.2900
MARS	**3.4728**	**1.8635**	**0.6438**	**0.4029**
Panel C:h = 10				
ARX	5.2503	2.2914	0.9083	0.1209
HAR	4.9636	2.2279	0.8129	0.1449
HARX	5.2942	2.3009	0.9163	0.0907
AHARX	5.0587	2.2491	0.9254	0.1296
HAR-RS-I	4.9398	2.2226	0.8094	0.1490
HAR-RS-II	4.8730	2.2075	0.8086	0.1607
HAR-SJ-I	4.9706	2.2295	0.8102	0.1440
HAR-SJ-II	4.9406	2.2227	0.8129	0.1493
FHAR	5.0929	2.2567	0.8277	0.1432
MARS	**4.5876**	**2.1419**	**0.7118**	**0.2083**
Panel D:h = 22				
ARX	6.1924	2.4885	1.2156	0.0633
HAR	5.1760	2.2751	0.9664	0.0704
HARX	5.9424	2.4377	1.1878	0.0645
AHARX	6.1417	2.4782	1.2123	0.0630
HAR-RS-I	5.0244	2.2415	0.9627	0.0978
HAR-RS-II	4.9523	2.2254	0.9626	0.1105
HAR-SJ-I	5.0715	2.2520	0.9640	0.0893
HAR-SJ-II	5.2210	2.2849	0.9675	0.0628
FHAR	5.1661	2.2729	0.9676	0.0713
MARS	**4.7338**	**2.1757**	**0.7977**	**0.1155**

The sample period for the NASDAQ 100 index spans from January 7, 1998 to March 24, 2017 (a total of 1,810,756 observations). We use a rolling window of 1000 observations to estimate the coefficients of all the models, and evaluate the out-of-sample forecast performance at four horizons ($h = 1, h = 5, h = 10$, and $h = 22$). Each panel in Table 4 corresponds to a specific forecast horizon, which varies from 1 to 22 days. Bold numbers indicate the best performing model by each criterion at each forecast horizon.

marginally improves. When the forecast horizon increases to 5, 10, and 22 days ahead, the improvement of the MARS method becomes unwavering. Also, it is meaningful to examine the results of HAR-RS-II, which is the base model for the model averaging method. The p-values for the HAR-RS-II indicate that we cannot reject the null hypothesis of equal performance of the HAR and HAR-RS-II models at $h = 5$, $h = 10$, and $h = 22$.

4.2.2. Dynamics of the model weights

Using the rolling window forecasts and the model averaging MARS method for the NASDAQ 100 index enables us to obtain the dynamics of the empirical model weights over the sample period. To describe the dynamics of the weights across time for varying forecast horizons, we consider the following models within the candidate model set:

$$\begin{aligned}\text{Model 1}: logRV_{t+h} &= \beta_0 + \beta_1 logRV_t^{(1)} \cdot \mathbb{I}(r_t < 0) + \beta_d^+ logRS_t^+ + \beta_d^- logRS_t^- \\ &+ \beta_w logRV_t^{(5)} + \beta_m logRV_t^{(22)} + \varepsilon_{t+h},\end{aligned}$$

Table 5
Giacomini-white test for the mean absolute forecast error.

Model	ARX	HAR	HARX	AHARX	HAR-RS -I	HAR-RS -II	HAR-SJ -I	HAR-SJ -II	FHAR
Panel A:h = 1									
HAR	0.0000	-	-	-	-	-	-	-	-
HARX	0.0001	0.0000	-	-	-	-	-	-	-
AHARX	0.0000	0.0000	0.0388	-	-	-	-	-	-
HAR-RS-I	0.0000	0.0035	0.0000	0.0000	-	-	-	-	-
HAR-RS-II	0.0000	0.0106	0.0000	0.0000	0.1799	-	-	-	-
HAR-SJ-I	0.0000	0.0029	0.0000	0.0000	0.6459	0.2284	-	-	-
HAR-SJ-II	0.0000	0.1864	0.0000	0.0000	0.2039	0.0516	0.0824	-	-
FHAR	0.0000	0.2871	0.0182	0.0000	0.0000	0.0000	0.0000	0.0000	-
MARS	**0.0000**	**0.0046**	**0.0000**	**0.0000**	0.0799	0.1364	0.0828	**0.0211**	**0.0000**
Panel B:h = 5									
HAR	0.0000	-	-	-	-	-	-	-	-
HARX	0.2059	0.0000	-	-	-	-	-	-	-
AHARX	0.4229	0.0000	0.1623	-	-	-	-	-	-
HAR-RS-I	0.0000	0.0914	0.0000	0.0000	-	-	-	-	-
HAR-RS-II	0.0000	0.0961	0.0000	0.0000	0.4301	-	-	-	-
HAR-SJ-I	0.0000	0.1656	0.0000	0.0000	0.0415	0.2386	-	-	-
HAR-SJ-II	0.0000	0.0162	0.0000	0.0000	0.0139	0.9437	0.0017	-	-
FHAR	0.0000	0.1772	0.0000	0.0000	0.1099	0.0898	0.1562	0.0991	-
MARS	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0001**	**0.0000**	**0.0000**	**0.0000**
Panel C:h = 10									
HAR	0.0117	-	-	-	-	-	-	-	-
HARX	0.7789	0.0000	-	-	-	-	-	-	-
AHARX	0.5030	0.0000	0.4571	-	-	-	-	-	-
HAR-RS-I	0.0088	0.1799	0.0000	0.0000	-	-	-	-	-
HAR-RS-II	0.0057	0.1993	0.0000	0.0000	0.7743	-	-	-	-
HAR-SJ-I	0.0091	0.2696	0.0000	0.0000	0.4639	0.5437	-	-	-
HAR-SJ-II	0.0081	0.9994	0.0000	0.0000	0.2873	0.0943	0.3682	-	-
FHAR	0.0000	0.8548	0.0001	0.0000	0.4756	0.7564	0.4547	0.8945	-
MARS	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
Panel D:h = 22									
HAR	0.0000	-	-	-	-	-	-	-	-
HARX	0.3771	0.0001	-	-	-	-	-	-	-
AHARX	0.8082	0.0000	0.1953	-	-	-	-	-	-
HAR-RS-I	0.0000	0.3286	0.0001	0.0000	-	-	-	-	-
HAR-RS-II	0.0000	0.4566	0.0001	0.0000	0.9757	-	-	-	-
HAR-SJ-I	0.0000	0.4806	0.0001	0.0000	0.1680	0.6390	-	-	-
HAR-SJ-II	0.0000	0.7267	0.0001	0.0000	0.0838	0.2629	0.0804	-	-
FHAR	0.0000	0.5442	0.0000	0.0000	0.6876	0.8754	0.2545	0.6544	-
MARS	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**

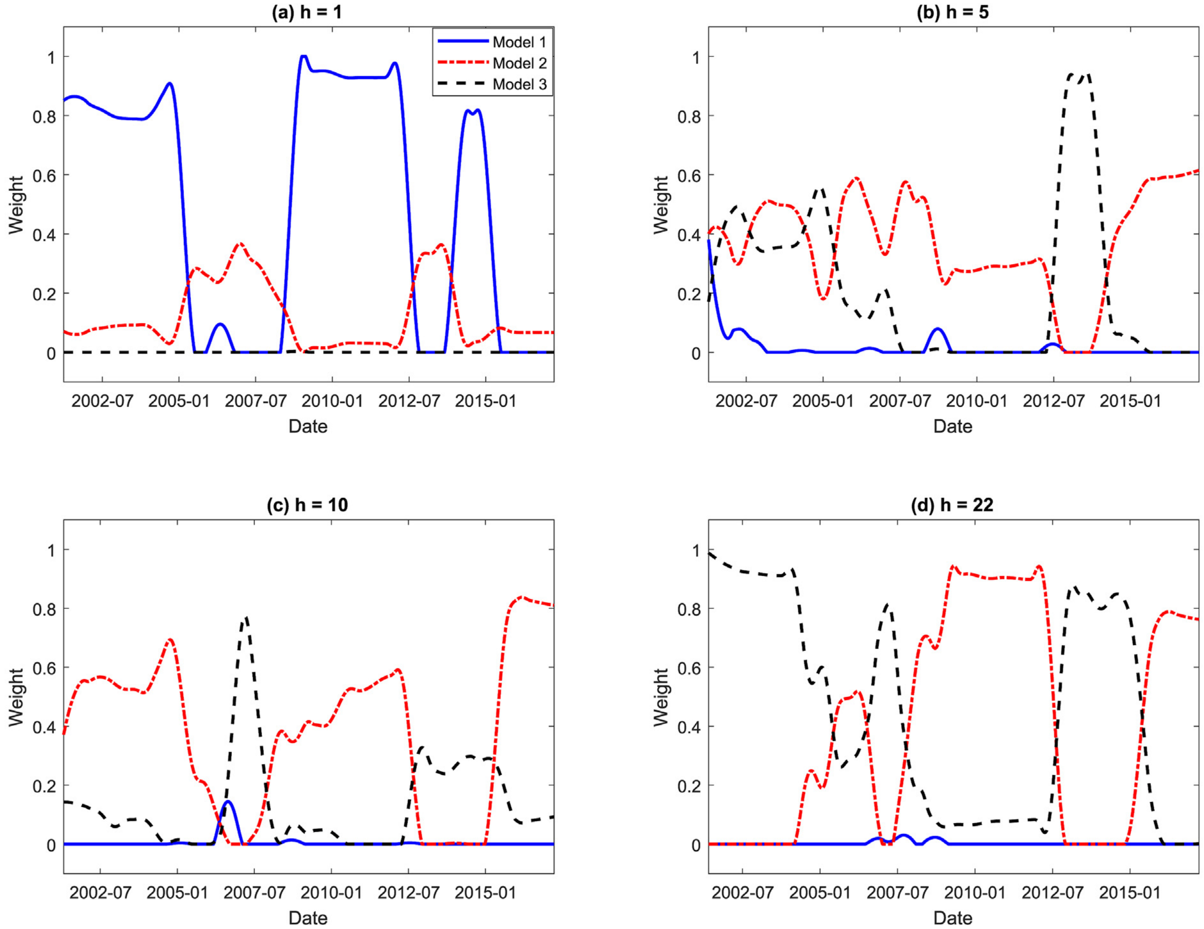
The sample period for the NASDAQ 100 index runs from January 7, 1998 to March 24, 2017, altogether 1,810,756 observations. We use a rolling window of 1000 observations to estimate all the models, and evaluate the out-of-sample forecast performance in the remaining series. The modified Giacomini-White test (Giacomini & White, 2006) is implemented to test the null hypothesis that the *row method* (in vertical headings) performs equally well to the *column method* (in horizontal headings) in terms of the absolute forecast error. Corresponding p values for a number of forecasting horizons ($h = 1, 5, 10, 22$) are reported in Panels A to D of Table 5, respectively.

$$\text{Model 2}: logRV_{t+h} = \beta_0 + \beta_1 logRV_t^{(1)} \cdot \mathbb{I}(r_t < 0) + \beta_2 logRV_t^{(2)} + \beta_3 logRV_t^{(3)} + \beta_{10} logRV_t^{(10)} + \varepsilon_{t+h},$$

$$\text{Model 3}: logRV_{t+h} = \beta_0 + \beta_1 logRV_t^{(1)} \cdot \mathbb{I}(r_t < 0) + \beta_2 logRV_t^{(2)} + \beta_{10} logRV_t^{(10)} + \varepsilon_{t+h},$$

where Model 1 is the base model-HAR-RS-II model; Model 2 contains the leverage term $RV_t^{(1)} \cdot \mathbb{I}(r_t < 0)$ and the HAR components with the lag index of $[2, 3, 10]$; and Model 3 is nested within Model 2, excluding the HAR component $logRV_t^{(3)}$. The reason for selecting the above three models is that they generally take up higher model weights, relative to other potential models across all forecast horizons.

The dynamics of weights for the three models at different forecast horizons are depicted in Fig. 1. Panels (a) to (d) represent the results for various forecast horizons $h = 1, 5, 10$, and 22, respectively. Within each panel, the solid line, dash-solid line, and dashed line separately capture the estimated weights (the vertical axis) for the above three models over different forecast periods (the horizontal axis). Only in the case of $h = 1$, Model 1 with the conventional lag structure $(1, 5, 22)$, completely



Note: This figure shows the dynamics of weights for the three models on the NASDAQ 100 index at different forecast horizons. Panels (a) to (d) represent the results for various forecast horizons $h = 1, 5, 10, 22$, respectively. Within each panel, the solid line, dash-solid line, and dashed line separately capture the estimated weights (the vertical axis) for the above three models over different forecast periods (the horizontal axis).

Fig. 1. Dynamics of Weights for the 3 Selected Models Over Time.
Note: This figure shows the dynamics of weights for the three models on the NASDAQ 100 index at different forecast horizons. Panels (a) to (d) represent the results for various forecast horizons $h = 1, 5, 10, 22$, respectively. Within each panel, the solid line, dash-solid line, and dashed line separately capture the estimated weights (the vertical axis) for the above three models over different forecast periods (the horizontal axis).

dominates the other two models in receiving higher weights (greater than 0.8) over most of the forecast period. This outcome does not hold during the 2007–2009 crisis. For the remaining forecast horizons, Model 1 is entirely surpassed by Models 2 and 3 in terms of the estimated model weights, and the importance of Models 2 and 3 alternates over the sample period.

We choose HAR-RS-II as the base model because, as demonstrated in Tables 4 and 5, HAR-RS-II performs better than other methods with no regard for model averaging. Nevertheless, being the best performer does not necessarily imply that HAR-RS-II will receive the highest model weights in all cases. In fact, evaluated by the estimated weights in Fig. 1, HAR-RS-II is surpassed by Models 2 and 3 when the forecast horizon exceeds $h = 1$.[21] We then take a closer look at Models 2 and 3, which reveals some interesting findings. First, neither of them belongs to the conventional HAR-type model with the lag index vector $[1, 5, 22]$. Second, they are only candidate models generated during the model averaging process. Finally, both models exclude the semivariance terms RS^{+} and RS^{-}. This implies that in our case, the joint effects of RS^{+} and RS^{-} is ignorable or at least one of them is redundant for long-horizon forecasts ($h = 5, 10$, and 22).[22] Moreover, the finding that no single model occupies the highest model weight in every case reinforces our belief that it is imperative to acknowledge model uncertainty.

[21] Selecting the best base model is beyond the scope of this paper and left for future research.

[22] This result coincides with the empirical results in Patton and Sheppard (2015). In their estimation, Patton and Sheppard (2015) found that $RS^{\rightarrow}$ provides more explanatory power than RS^{+} in almost all cases.

Table 6
The winning rates of the MARS method in various situations.

	$h=1$	$h=5$	$h=10$	$h=22$
MSFE	73%	79%	81%	79%
SDFE	73%	79%	81%	79%
MAFE	74%	82%	83%	84%
Pseudo R^2	74%	80%	82%	79%
GW Test at 10%	64%	73%	77%	71%
GW Test at 5%	34%	70%	71%	70%

The upper panel reports the winning rate of the MARS method across all individual stocks, that is, by specific criterion at a certain forecast horizon, the percentage of the MARS method outperforming the alternatives for all the individual stocks. The lower panel contains the percentages of the MARS method significantly improving over the alternatives across all the individual stocks, which are generated from the p-values of the GW test statistic. Results based on the 10% and 5% levels of significance are presented in the last two rows, respectively.

The result here also complements the conclusions in Section 4.2.1 by illustrating the importance of a flexible lag structure under the HAR framework.

4.3. Individual stocks

In this section, we extend our analysis to the 104 individual stocks, which are components of the NASDAQ 100 index. Their realized variances are also estimated following Equation (1). Note that the forecasting exercise detailed in Section 4.2.1 is repeated on each of the individual stocks. For each stock, the base model for the model averaging method is selected from one of the four HAR-RS type models, which can vary across individual stocks.

Since it would be exhaustive to present the out-of-sample results separately for the 104 individual stocks, we summarize the winning rates of the model averaging method (i.e., the MARS method) in Table 6,[23] in which the upper panel reports the winning rate of the MARS model across all individual stocks, that is, by specific criterion at a certain forecast horizon, the percentage of the MARS method outperforming the alternatives across all the individual stocks. The winning rates of the MARS method are higher than 70% across all of the forecast horizons, as well as by each forecast criterion. Moreover, we note that, as the forecast horizon expands from 1 day to 5 days—10 days and 22 days ahead, our MARS method is more likely to outperform the alternatives. The highest winning rate (84%) for the MARS method occurs at $h=22$ and the MAFE acts as the criterion, while the lowest winning rate for the MARS method occurs at $h=1$, and the forecasts are assessed using the MSFE and SDFE.

The lower panel of Table 6 contains the percentages of the MARS method significantly improving over the alternatives across all of the individual stocks, generated from the p-values of the GW test statistic. The results based on the 10% and 5% levels of significance are presented in the last two rows of Table 6, respectively. At the 10% level, we conclude that for more than 60% of the stocks, our MARS method can significantly improve the forecasting accuracy over the alternative models. The winning rates decrease once we set the confidence level at the 5% level. At $h=1$, the winning rate drops sharply to 34%. However, as h increases to 5 days and beyond, the winning rates are approximately the same level as their counterparts at the 10% level. In fact, in regard to $h=22$, the winning rates at the 10% and 5% levels are quite close.

5. Conclusions

In this paper, we propose using the least squares model averaging method to allow for a more general lag structure under the HAR framework, instead of restricting it to $(1,5,22)$. Specially, we estimate the HAR-RS type models in Patton and Sheppard (2015) with the least squares model averaging method and consider constructing the potential model set with a full permutation of all of the possible lags and a maximum lag order of 22. The MARS embedded model is data-driven, as the empirical weights on potential models with different lag combinations change with various volatility series and different forecast horizons. We prove that the MARS estimator is asymptotically optimal in the sense of achieving the lowest possible mean squared forecast error.

In an empirical application to high-frequency data from the NASDAQ 100 index and the associated 104 individual stocks, we show that the MARS method can generally outperform the alternatives under various forecast criteria as well as across all forecast horizons ($h=1,5,10,22$). Specifically, we apply the GW test to examine the statistical significance of the improvement made by the MARS method. We find that the MARS method performs significantly better than most of HAR-type comparison models at a 10% confidence level under the forecast horizon $h=1$, yet the improvement by the MARS model over the HAR-RS-II model is marginally significant. Conversely, when the forecast horizon increases to 5, 10, and 22 days ahead, the improvement by the MARS method relative to other models is highly significant, even at the 1% confidence level. The above findings suggest that considering the model specification uncertainty under the HAR framework is

[23] Though the details are not reported here, they are available upon request.

meaningful because the model averaging method-based estimation can generate more accurate forecasts than individual models in both statistical and economic senses.

Acknowledgments

This paper is partially supported by the National Natural Science Foundation of China (71701175, 71522004, 11471324, 71631008, and 71642006), the Ministry of Education of the People's Republic of China Project of Humanities and Social Sciences (17YJC790174 and 17YJC910011), the Natural Science Foundation of Fujian Province of China (2018J01116), the Fundamental Research Funds for the Central Universities China (20720171002, 20720171076, 20720181050, and 20720181004), and the Educational and Scientific Research Program for Young and Middle-aged Instructors of Fujian Province (JAS170018). We wish to thank Steven F. Lehrer, Bruno Solnik, Lu Zhang, the seminar participants at the Young Econometricians in the Asian-Pacific (YEAP) region 2017 annual conference, the Chinese Academy of Sciences, Renmin University, and Xiamen University for helpful comments and suggestions. Tian Xie is grateful to Jiarong Chen for excellent research assistance. Any errors are our own.

Conflicts of interest

The authors declare no conflict of interest.

APPENDIX

A Proof of Theorem 1

Let $\mathrm{MARS}^*(\boldsymbol{w}) = \mathrm{MARS}(\boldsymbol{w}) - \|\varepsilon\|^2 + 2\varepsilon^\top \boldsymbol{\mu}$. It is known that

$$\widehat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}\in\mathscr{H}} \mathrm{MARS}^*(\boldsymbol{w}).$$

Using Lemma 1 in Zhang (2010), we need to verify the below equations to prove Theorem 1,

$$\sup_{\boldsymbol{w}\in\mathscr{H}} \frac{L_T(\boldsymbol{w}) - \tilde{L}_T(\boldsymbol{w})}{\tilde{L}_T(\boldsymbol{w})} \to 0 \tag{A1}$$

and

$$\sup_{\boldsymbol{w}\in\mathscr{H}} \frac{\mathrm{MARS}^*(\boldsymbol{w}) - L_T(\boldsymbol{w})}{\tilde{L}_T(\boldsymbol{w})} \to 0. \tag{A2}$$

By Conditions (C.1) and (C.3), we obtain

$$\begin{aligned}
&\sup_{\boldsymbol{w}\in\mathscr{H}} \left| L_T(\boldsymbol{w}) - \tilde{L}_T(\boldsymbol{w}) \right| \\
&= \sup_{\boldsymbol{w}\in\mathscr{H}} \left| \left\| \boldsymbol{X}\widehat{\beta}(\boldsymbol{w}) - \boldsymbol{\mu} \right\|^2 - \left\| \boldsymbol{X}\tilde{\beta}(\boldsymbol{w}) - \boldsymbol{\mu} \right\|^2 \right| \\
&= \sup_{\boldsymbol{w}\in\mathscr{H}} \left| \left\{ \boldsymbol{X}\widehat{\beta}(\boldsymbol{w}) - \boldsymbol{X}\tilde{\beta}(\boldsymbol{w}) \right\}^\top \left\{ \boldsymbol{X}\widehat{\beta}(\boldsymbol{w}) + \boldsymbol{X}\tilde{\beta}(\boldsymbol{w}) - 2\boldsymbol{\mu} \right\} \right| \\
&= \sup_{\boldsymbol{w}\in\mathscr{H}} \left| \sum_{m=1}^{M} \sum_{l=1}^{M} w_m w_l \left\{ \boldsymbol{X\Pi}^{m\top} \widehat{\beta}^m - \boldsymbol{X\Pi}^{m\top} \tilde{\beta}^m \right\}^\top \left\{ \boldsymbol{X\Pi}^{l\top} \widehat{\beta}^l + \boldsymbol{X\Pi}^{l\top} \tilde{\beta}^l - 2\boldsymbol{\mu} \right\} \right| \\
&\leq \sum_{m=1}^{M} \sum_{l=1}^{M} \left| \left\{ \boldsymbol{X\Pi}^{m\top} \widehat{\beta}^m - \boldsymbol{X\Pi}^{m\top} \tilde{\beta}^m \right\}^\top \left\{ \boldsymbol{X\Pi}^{l\top} \widehat{\beta}^l + \boldsymbol{X\Pi}^{l\top} \tilde{\beta}^l - 2\boldsymbol{\mu} \right\} \right| \\
&= O_p\left(T^{1/2} M^2 \right),
\end{aligned} \tag{A3}$$

where the last equality uses Conditions (C.1) and (C.3). Also, by Conditions (C.1) and (C.3),

$$
\begin{aligned}
&\sup_{\boldsymbol{w}\in\mathscr{H}}\left|\mathrm{MARS}^{*}(\boldsymbol{w})-L_T(\boldsymbol{w})\right| \\
&=\sup_{\boldsymbol{w}\in\mathscr{H}}\left|\|\boldsymbol{y}-\widehat{\boldsymbol{\mu}}(\boldsymbol{w})\|^2\frac{T+k(\boldsymbol{w})}{T-k(\boldsymbol{w})}-\|\boldsymbol{\varepsilon}\|^2+2\boldsymbol{\varepsilon}^{\top}\boldsymbol{\mu}-\left\|\boldsymbol{X}\widehat{\beta}(\boldsymbol{w})-\boldsymbol{\mu}\right\|^2\right| \\
&=\sup_{\boldsymbol{w}\in\mathscr{H}}\left|\left\|\boldsymbol{X}\widehat{\beta}(\boldsymbol{w})-\boldsymbol{y}\right\|^2-\left\|\boldsymbol{X}\widehat{\beta}(\boldsymbol{w})-\boldsymbol{\mu}\right\|^2-\|\boldsymbol{\varepsilon}\|^2+2\boldsymbol{\varepsilon}^{\top}\boldsymbol{\mu}\|+\boldsymbol{y}-\widehat{\boldsymbol{\mu}}(\boldsymbol{w})\|^2\frac{2k(\boldsymbol{w})}{T-k(\boldsymbol{w})}\right| \\
&=\sup_{\boldsymbol{w}\in\mathscr{H}}\left|-2\boldsymbol{\varepsilon}^{\top}\boldsymbol{X}\widehat{\beta}(\boldsymbol{w})+\|\boldsymbol{y}-\widehat{\boldsymbol{\mu}}(\boldsymbol{w})\|^2\frac{2k(\boldsymbol{w})}{T-k(\boldsymbol{w})}\right| \\
&=\sup_{\boldsymbol{w}\in\mathscr{H}}\left|-2\boldsymbol{\varepsilon}^{\top}\boldsymbol{X}\tilde{\beta}(\boldsymbol{w})+2\boldsymbol{\varepsilon}^{\top}\boldsymbol{X}\left\{\widehat{\beta}(\boldsymbol{w})-\tilde{\beta}(\boldsymbol{w})\right\}+\|\boldsymbol{y}-\widehat{\boldsymbol{\mu}}(\boldsymbol{w})\|^2\frac{2k(\boldsymbol{w})}{T-k(\boldsymbol{w})}\right| \\
&=\sup_{\boldsymbol{w}\in\mathscr{H}}\left|2\sum_{m=1}^{M}w_m\boldsymbol{\varepsilon}^{\top}\boldsymbol{X}\boldsymbol{\Pi}^{m\top}\tilde{\beta}^{m}\right|+\sup_{\boldsymbol{w}\in\mathscr{H}}\left|2\sum_{m=1}^{M}w_m\boldsymbol{\varepsilon}^{\top}\boldsymbol{X}\left(\boldsymbol{\Pi}^{m\top}\widehat{\beta}^{m}-\boldsymbol{\Pi}^{m\top}\tilde{\beta}^{m}\right)\right| \\
&\quad+\sup_{\boldsymbol{w}\in\mathscr{H}}\|\boldsymbol{y}-\widehat{\boldsymbol{\mu}}(\boldsymbol{w})\|^2\frac{2k(\boldsymbol{w})}{T-k(\boldsymbol{w})} \\
&=O_p\left(T^{1/2}M\right)+O_p\left(\max_{m\in\{1,\dots,M\}}k_m\right),
\end{aligned}
\tag{A4}
$$

where the last step is from Conditions (C.1) and (C.3). Combining the above two results and Condition (C.2), we reach (A1)-(A2). This completes the proof.

B Details on the NASDAQ 100 Stocks

In the empirical exercise, we include the individual stocks that composed of the NASDAQ 100 index to examine the robustness of our proposed MARS approach. The NASDAQ 100 Index includes 104 of the largest domestic and international non-financial companies listed on the NASDAQ Stock Market based on market capitalization. The Index reflects companies across major industry groups including computer hardware and software, telecommunications, retail/wholesale trade, and biotechnology. Their NASDAQ tickers, associated company names, and industries are listed in the second, third, and fourth columns of Table A1, respectively.

Table A1

Descriptions of the 104 Individual Stocks

No.	Ticker Symbol	Company Name	GICS Sector
1	ATVI	Activision Blizzard	Information Technology
2	ADBE	Adobe Systems Incorporated	Information Technology
3	ALXN	Alexion Pharmaceuticals	Health Care
4	ALGN	Align Technology, Inc	Health Care
5	GOOGL	Alphabet Inc. Class A	Information Technology
6	GOOG	Alphabet Inc. Class C	Information Technology
7	AMZN	Amazon.com, Inc	Consumer Discretionary
8	AAL	American Airlines Group	Industrials
9	AMGN	Amgen Inc	Health Care
10	ADI	Analog Devices	Information Technology
11	AAPL	Apple Inc	Information Technology
12	AMAT	Applied Materials, Inc	Information Technology
13	ASML	ASML Holding	Information Technology
14	ADSK	Autodesk, Inc	Information Technology
15	ADP	Automatic Data Processing, Inc	Information Technology
16	BIDU	Baidu.com, Inc	Information Technology
17	BIIB	Biogen, Inc	Health Care
18	BMRN	BioMarin Pharmaceutical, Inc	Health Care
19	AVGO	Broadcom Limited	Information Technology
20	CA	CA Technologies	Information Technology
21	CDNS	Cadence Design Systems, Inc.	Information Technology
22	CELG	Celgene Corporation	Health Care
23	CERN	Cerner Corporation	Health Care
24	CHTR	Charter Communications, Inc	Consumer Discretionary

(continued on next page)

Table A1 (*continued*)

No.	Ticker Symbol	Company Name	GICS Sector
25	CHKP	Check Point Software Technologies Ltd.	Information Technology
26	CTAS	Cintas Corporation	Industrials
27	CSCO	Cisco Systems, Inc	Information Technology
28	CTXS	Citrix Systems, Inc	Information Technology
29	CTSH	Cognizant Technology Solutions Corporation	Information Technology
30	CMCSA	Comcast Corporation	Consumer Discretionary
31	COST	Costco Wholesale Corporation	Consumer Staples
32	CSX	CSX Corporation	Industrials
33	CTRP	CTrip International	Consumer Discretionary
34	XRAY	Dentsply Sirona	Health Care
35	DISH	Dish Network, Inc	Consumer Discretionary
36	DLTR	Dollar Tree, Inc	Consumer Discretionary
37	EBAY	eBay Inc	Information Technology
38	EA	Electronic Arts	Information Technology
39	EXPE	Expedia, Inc	Consumer Discretionary
40	ESRX	Express Scripts, Inc	Health Care
41	FB	Facebook, Inc	Information Technology
42	FAST	Fastenal Company	Industrials
43	FISV	Fiserv, Inc	Information Technology
44	GILD	Gilead Sciences, Inc	Health Care
45	HAS	Hasbro, Inc	Consumer Discretionary
46	HSIC	Henry Schein, Inc	Health Care
47	HOLX	Hologic, Inc	Health Care
48	IDXX	IDEXX Laboratories, Inc	Health Care
49	ILMN	Illumina, Inc	Health Care
50	INCY	Incyte Corporation	Health Care
51	INTC	Intel Corporation	Information Technology
52	INTU	Intuit, Inc.	Information Technology
No.	Ticker Symbol	Company Name	GICS Sector
53	ISRG	Intuitive Surgical Inc.	Health Care
54	JBHT	J.B. Hunt Transport Services, Inc.	Industrials
55	JD	JD.com	Consumer Discretionary
56	KLAC	KLA-Tencor Corporation	Information Technology
57	LRCX	Lam Research, Inc.	Information Technology
58	LBTYA	Liberty Global plc Ordinary A	Consumer Discretionary
59	LBTYK	Liberty Global plc Ordinary C	Consumer Discretionary
60	LVNTA	Liberty Interactive	Consumer Discretionary
61	QVCA	Liberty Interactive	Consumer Discretionary
62	MAR	Marriott International, Inc.	Consumer Discretionary
63	MXIM	Maxim Integrated Products	Information Technology
64	MELI	MercadoLibre	Information Technology
65	MCHP	Microchip Technology	Information Technology
66	MU	Micron Technology, Inc.	Information Technology
67	MSFT	Microsoft Corporation	Information Technology
68	MDLZ	Mondelẽz International	Consumer Staples
69	MNST	Monster Beverage	Consumer Staples
70	MYL	Mylan N.V.	Health Care
71	NTES	NetEase, Inc.	Information Technology
72	NFLX	Netflix	Information Technology
73	NVDA	NVIDIA Corporation	Information Technology
74	ORLY	O'Reilly Automotive, Inc.	Consumer Discretionary
75	PCAR	PACCAR Inc.	Industrials
76	PAYX	Paychex, Inc.	Information Technology
77	PYPL	PayPal Holdings, Inc.	Information Technology
78	QCOM	QUALCOMM Incorporated	Information Technology
79	REGN	Regeneron Pharmaceuticals	Health Care
80	ROST	Ross Stores Inc.	Consumer Discretionary
81	STX	Seagate Technology Holdings	Information Technology
82	SHPG	Shire plc	Health Care
83	SIRI	Sirius XM Radio, Inc.	Consumer Discretionary
84	SWKS	Skyworks Solutions, Inc.	Information Technology
85	SBUX	Starbucks Corporation	Consumer Discretionary
86	SYMC	Symantec Corporation	Information Technology
87	SNPS	Synopsys, Inc.	Information Technology
88	TMUS	T-Mobile US	Telecommunication Services
89	TTWO	Take-Two Interactive, Inc.	Information Technology
90	TSLA	Tesla, Inc.	Consumer Discretionary
91	TXN	Texas Instruments, Inc.	Information Technology
92	KHC	The Kraft Heinz Company	Consumer Staples

Table A1 (*continued*)

No.	Ticker Symbol	Company Name	GICS Sector
93	PCLN	The Priceline Group	Consumer Discretionary
94	FOXA	Twenty-First Century Fox Class A	Consumer Discretionary
95	FOX	Twenty-First Century Fox Class B	Consumer Discretionary
96	ULTA	Ulta Beauty	Consumer Discretionary
97	VRSK	Verisk Analytics	Industrials
98	VRTX	Vertex Pharmaceuticals	Health Care
99	VOD	Vodafone Group, plc.	Telecommunication Services
100	WBA	Walgreens Boots Alliance	Consumer Staples
101	WDC	Western Digital	Information Technology
102	WDAY	Workday, Inc.	Information Technology
103	WYNN	Wynn Resorts	Consumer Discretionary
104	XLNX	Xilinx, Inc.	Information Technology

C Discussion on the Exogenous Variable $\boldsymbol{z}_t$

As defined in Equations (4) and (5), HARX and AHARX models include exogenous variables $\boldsymbol{z}_t$. In our exercises, the exogenous variables $\boldsymbol{z}_t$ consist of the following seven variables:

(i) NASDAQ 100: The logarithm of the NASDAQ 100 index;
(ii) VOL: The logarithm of the S&P 500 vol;
(iii) OIL: The logarithm of the one-month crude oil future contract index;
(iv) USDI: The logarithm of the trade-weighted average of the foreign exchange value of the US dollar index;
(v) CS: Credit Spread, the excess yield of the Moody's seasoned Baa corporate bond over the Moody's seasoned Aaa corporate bond;
(vi) TS: Term Spread, the difference between the 10-Year and 3-month treasury constant maturity rates;
(vii) FFD: Federal Fund Deviation, the difference between the effective and target Federal Fund rates.

Descriptive statistics of $\boldsymbol{z}_t$ are presented in Table A2. We report the logarithm of the NASDAQ 100 index, NASDAQ 100 vol, one-month crude oil future contract index, and trade-weighted average of the foreign exchange value of the US dollar index, instead of their observed values.

Table A2

Descriptive Statistics for the Logarithm of Exogenous Variables

Sample Statistics	NASDAQ 100	VOL	Oil	USDI	CS	TS	FFD
Mean	6.896	20.924	3.633	4.461	0.962	1.839	0.014
Median	7.041	21.085	3.492	4.466	0.880	1.950	0.000
Min	5.689	17.862	2.372	4.220	0.500	−0.950	−1.810
Max	7.771	23.162	4.982	4.728	3.500	3.870	3.640
Standard Deviation	0.524	1.144	0.654	0.114	0.403	1.128	0.180
Skewness	−0.631	−0.381	0.192	0.056	3.120	−0.250	2.848
Kurtosis	2.368	1.881	1.645	2.354	16.403	2.083	56.679
Jarque-Bera Test	0.001	0.001	0.001	0.001	0.001	0.001	0.001
ADF Test	0.988	0.599	0.611	0.650	0.310	0.235	0.001

As indicated by the *p*-values of the ADF test in Table A2, all the series, except the federal fund deviation, are nonstationary. For these nonstationary data series, we use their log differenced values in the exercises. Although not reported here, all series after log transformation are stationary. Moreover, the results of the Jarque-Bera test imply that none of the above series is normally distributed.

D Comparison Between Various Types of MARS

This section extends the empirical exercise in Section 4.2.1. We further examine the forecast performance of the MARS methods based on various HAR-RS model specifications, where the four HAR-RS models described in Equations (6)–(9) are individually used as the base model for constructing the potential model set. They are denoted as MARS_1 to MARS_4 in Table A3, respectively.

Similar to the experiment in Section 4.2.1, we conduct a rolling-window exercise for the RV of the NASDAQ 100 index on models $MARS_1$ to $MARS_4$. The results are reported in Table A3. The first column lists the four MARS models considered here. Columns 2–5 correspond to the results of forecast criteria, the MSFE, SDFE, MAFE, and Pseudo R^2, respectively. Panels A to D represent different forecast horizons varying from one day ahead to 22 days ahead. Bold numbers indicate the best performing model by each criterion at each forecast horizon. It is obvious that the $MARS_2$ model, which is based on HAR-RS-II, has the best performance in all cases. This validates our choice of using the HAR-RS-II as the base model for the MARS method in Section 4.2.1.

Table A3

Out-Of-Sample Comparison of Different MARS Models for the RV of the NASDAQ 100 Index

Model	MSFE	SDFE	MAFE	Pseudo R^2
Panel A:h = 1				
$MARS_1$	2.4552	1.5669	0.5078	0.5789
$MARS_2$	**1.8640**	**1.3653**	**0.4973**	**0.6809**
$MARS_3$	2.4918	1.5785	0.5110	0.5714
$MARS_4$	2.1951	1.4816	0.5251	0.6232
Panel B:h = 5				
$MARS_1$	3.7404	1.9340	0.6677	0.3535
$MARS_2$	**3.4728**	**1.8635**	**0.6438**	**0.4029**
$MARS_3$	3.6856	1.9198	0.6669	0.3619
$MARS_4$	3.4951	1.8695	0.6628	0.3959
Panel C:h = 10				
$MARS_1$	5.0245	2.2415	0.7181	0.1994
$MARS_2$	**4.5876**	**2.1419**	**0.7118**	**0.2083**
$MARS_3$	4.9945	2.2348	0.7168	0.2041
$MARS_4$	5.0153	2.2395	0.7173	0.2012
Panel D:h = 22				
$MARS_1$	5.1107	2.2607	0.8095	0.0763
$MARS_2$	**4.7338**	**2.1757**	**0.7977**	**0.1155**
$MARS_3$	5.0440	2.2459	0.8115	0.0882
$MARS_4$	5.0338	2.2436	0.8080	0.0901

The sample period for the NASDAQ 100 index spans from January 7, 1998 to March 24, 2017 (a total of 1,810,756 observations). We use a rolling window of 1000 observations to estimate various MARS models and evaluate the out-of-sample forecast performance at four horizons ($h = 1, h = 5, h = 10$, and $h = 22$). Each panel in Table A3 corresponds to a specific forecast horizon, which varies from 1 to 22 days. Bold numbers indicate the best performing model by each criterion at each forecast horizon.

References

Aït-Sahalia, Y., & Yu, J. (2009). High frequency market microstructure noise estimates and liquidity measures. *Annals of Applied Statistics, 3*(1), 422–457.

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

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

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001b). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association, 96*(453), 42–55.

Ando, T., & Li, K.-C. (2014). A model-averaging approach for high-dimensional regression. *Journal of the American Statistical Association, 109*(505), 254–265.

Audrino, F., Camponovo, L., & Roth, C. (2015). *Testing the lag structure of assets' realized volatility dynamics. Economics Working Paper Series 1501*. University of St. Gallen, School of Economics and Political Science. downloaded on from 15 March 2018 https://ideas.repec.org/p/usg/econwp/201501.html.

Audrino, F., Huang, C., & Okhrin, O. (2016). *Flexible har model for realized volatility. Working paper, St. Gallen*. downloaded on from 24 March 2018 https://www.alexandria.unisg.ch/248495/.

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

Becker, R., Clements, A. E., & White, S. I. (2007). Does implied volatility provide any information beyond that captured in model-based volatility forecasts? *Journal of Banking & Finance, 31*(8), 2535–2549.

Bollerslev, T., Litvinova, J., & Tauchen, G. (2006). Leverage and volatility feedback effects in high-frequency data. *Journal of Financial Econometrics, 4*(3), 353–384.

Bollerslev, T., & Zhou, H. (2006). Volatility puzzles: A simple framework for gauging return-volatility regressions. *Journal of Econometrics, 131*(1), 123–150.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (pp. 149–205). New York: Springer.

Chan, K., & Fong, W.-M. (2000). Trade size, order imbalance, and the volatility–volume relation. *Journal of Financial Economics, 57*(2), 247–273.

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

Corsi, F., Audrino, F., & Renò, R. (2012). Har modeling for realized volatility forecasting. In *Handbook of volatility models and their applications* (pp. 363–382). John Wiley & Sons, Inc.

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

Craioveanu, M., & Hillebrand, E. (2012). *Why it is ok to use the har-rv (1,5,21) model. Technical Report*. University of Central Missouri.

Dacorogna, M. M., Műller, U. A., Nagler, R. J., Olsen, R. B., & Pictet, O. V. (1993). A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *Journal of International Money and Finance, 12*(4), 413–438.

Fernandes, M., Medeiros, M. C., & Scharth, M. (2014). Modeling and predicting the cboe market volatility index. *Journal of Banking & Finance, 40*, 1–10.

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica, 74*(6), 1545–1578.
Hansen, B. E. (2008). Least-squares forecast averaging. *Journal of Econometrics, 146*(2), 342–350.
Lamoureux, C. G., & Lastrapes, W. D. (1990). Heteroskedasticity in stock return data: Volume versus garch effects. *The Journal of Finance, 45*(1), 221–229.
Lehrer, S. F., & Xie, T. (2017). Box office buzz: Does social media data steal the show from model uncertainty when forecasting for hollywood? *The Review of Economics and Statistics, 99*(5), 749–755.
Liu, C., & Maheu, J. M. (2009). Forecasting realized volatility: A bayesian model averaging approach. *Journal of Applied Econometrics, 24*(5), 709–733.
McAleer, M., & Medeiros, M. C. (2008). A multiple regime smooth transition heterogeneous rutoregressive model for long memory and asymmetries. *Journal of Econometrics, 147*(1), 104–119.
Müller, U. A., Dacorogna, M. M., Davé, R. D., Pictet, O. V., Olsen, R. B., & Ward, J. (1993). *Fractals and intrinsic time - a challenge to econometricians*. Technical report.
Patton, A. J., & Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *The Review of Economics and Statistics, 97*(3), 683–697.
Scharth, M., & Medeiros, M. (2009). Asymmetric effects and long memory in the volatility of dow jones stocks. *International Journal of Forecasting, 25*(2), 304–327.
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B, 58*, 267–288.
Wang, Y., Ma, F., Wei, Y., & Wu, C. (2016). Forecasting realized volatility in a changing world: A dynamic model averaging approach. *Journal of Banking & Finance, 64*, 136–149.
White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica, 50*, 1–25.
Xie, T. (2015). Prediction model averaging estimator. *Economics Letters, 131*, 5–8.
Xie, T. (2017). Heteroscedasticity-robust model screening: A useful toolkit for model averaging in big data analytics. *Economics Letters, 151*(C), 119–122.
Yuan, Z., & Yang, Y. (2005). Combining linear regression models: When and how? *Journal of the American Statistical Association, 100*(472), 1202–1214.
Zhang, X. (2010). *Model averaging and its applications*. Ph.D. Thesis, (1):Academy of Mathematics and Systems Science. Chinese Academy of Sciences.
Zhang, X., Lu, Z., & Zou, G. (2013a). Adaptively combined forecasting for discrete response time series. *Journal of Econometrics, 176*(1), 80–91.
Zhang, L., Mykland, P. A., & Aït-Sahalia, Y. (2005). A tale of two time scales. *Journal of the American Statistical Association, 100*(472), 1394–1411.
Zhang, X., Wan, A. T. K., & Zou, G. (2013b). Model averaging by jackknife criterion in models with dependent data. *Journal of Econometrics, 174*, 82–94.