Flexible HAR Model for Realized Volatility*

Francesco Audrino†, Chen Huang‡, Ostap Okhrin§

April 1, 2016

Abstract

The Heterogeneous Autoregressive (HAR) model is commonly used in modeling the dynamics of realized volatility. In this paper, we propose a flexible HAR$(1, \ldots, p)$ specification, employing the adaptive LASSO and its statistical inference theory to see whether the lag structure (1, 5, 22) implied from an economic point of view can be recovered by statistical methods.

Adaptive LASSO estimation and the subsequent hypothisis testing results show that there is no strong evidence that such a fixed lag structure can be exactly recovered by a flexible model. In terms of the out-of-sample forecasting, the proposed model slightly outperforms the classic specification and a superior predictive ability test shows that it cannot be significantly outperformed by any of the alternatives. We also apply the group LASSO and some related tests to check the validity of the classic HAR, which is rejected in most cases. The main reason for rejection might be the arrangement of groups, and a minor reason is the equality constraints on AR coefficients. This justifies our intention to use a flexible lag structure while still keeping the HAR frame. Finally, the time-varying behaviors show that when the market environment is not stable, the structure of (1, 5, 22) does not hold very well.

JEL classification: C12, C22, C51, C53

Keywords: Heterogeneous Autoregressive Model, Realized Volatility, Lag Structure, Adaptive LASSO, Hypothesis Testing

---

*Financial support from the Deutsche Forschungsgemeinschaft via CRC 649 "Economic Risk" and IRTG 1792 "High Dimensional Non Stationary Time Series", Humboldt-Universität zu Berlin, is gratefully acknowledged.

†Faculty of Mathematics and Statistics, University of St.Gallen, Bodanstrasse 6, 9000 St.Gallen, Switzerland. Email: francesco.audrino@unisg.ch

‡Humboldt-Universität zu Berlin, C.A.S.E. - Center for Applied Statistics and Economics, Spandauer Str. 1, 10178 Berlin, Germany. Email: chen.huang@hu-berlin.de

§Chair of Econometrics and Statistics, Faculty of Transportation, Dresden University of Technology, Würzburger Str. 35, 01187 Dresden, Germany. Email: ostap.okhrin@tu-dresden.de

1 Introduction

Clustering and long memory are basic characteristics in financial market volatility time series. Numerous papers focus on capturing these properties more accurately and providing a better forecasting performance. The most well known model is GARCH introduced by **?** and its series of extensions with the fractional integration GARCH (FIGARCH) model by **?** among them.

On the other hand, with increasing accessibility to high-frequency trading data, a great deal of research has been done to model and forecast the realized volatility constructed from high-frequency intra-day returns. ARFIMA type specifications have often been employed to model the time-varying dynamics of realized volatility, especially to capture its high persistency property; see for example **?**.

However, vast empirical analysis of financial data shows that volatilities over different time horizons have asymmetric interactions. Volatilities over longer time intervals have stronger influence on those at shorter time intervals than conversely; for example **?**. Such a "volatility cascade phenomenon" can be easily interpreted economically but it cannot be captured by standard volatility models. Based on the heterogeneous market hypothesis by **?**, a linear additive process with heterogeneous components called the Heterogeneous Autoregressive (HAR) model was proposed by **?**. Although it does not formally belong to the category of long memory models, empirically it is observed to be able to display apparent high persistency facts of financial time series. Moreover, due to its computational simplicity and excellent out-of-sample forecasting performance, the HAR model is commonly used in realized volatility applications. Several extensions of the HAR model have been recently proposed. They consider jump behaviors, leverage effects and others; see a survey by **?**.

The hierarchical structure assumed in the HAR model includes three partial components: short-term traders with daily or higher trading frequency, medium-term traders with weekly trading frequency, and long-term traders with monthly or lower trading frequency.

Therefore, the lag structure in the HAR is fixed as (1, 5, 22). But the suitability of such a specification is the topic of this study. **?** use all possible combinations of lags (chosen within the maximum lag of 250) for the last two terms in the additive model and compare their in-sample or out-of-sample fitting performance. Although their results support the classic HAR(1, 5, 22) assumption, the included components are still fixed at three and the cost of computation is enormous.

? find that the implied lag structure from an economic point of view cannot be recovered by the Least Absolute Shrinkage and Selection Operator (LASSO) technique introduced by **?** on statistical aspect, but they show equal forecasting performance. **?** employ the adaptive LASSO introduced by **?** as the variable selection method and make use of the inference theory of adaptive LASSO estimators in time series regression models developed by **?** to construct a conservative testing procedure to test the optimal lag structure of realized volatility dynamics. Since the realized volatility over longer time horizons (longer than daily) is defined as the sample average of daily realized volatility, the HAR model can in fact also be written as a constrained AR model. Both **?** and **?** work on the AR framework, as they apply the (adaptive) LASSO to select active AR lag terms. However, an arbitrary HAR model is a special AR model, but not every AR model can be converted back to a HAR model. In other words, the optimal AR lag structure after selection might not reflect the volatility cascade as the HAR model supposed to have. In particular, previous works did not check whether or not the coefficient constraints on AR terms implied by HAR models are satisfied. Therefore, this paper extends the work by **?** into the HAR framework. We follow a similar hypothesis testing procedure on the presence of false positives. But the results from LASSOing the HAR framework directly could be more comparable to the original HAR model. Furthermore, we compare the proposed flexible model with the fixed choice HAR(1, 5, 22), a three non-zero coefficient specification, namely HAR(a, b, c), an additive nonparametric model, and the HARQ model including realized quarticity (introduced by **?** recently) in terms of the in-sample fitting and out-of-sample forecasting performances. To test the validity of the classic HAR lag structure, we also employ the group LASSO to identify the active AR lags.

Group LASSO estimation implies that if one group is active, then all the variables in it will be active. Thus, if the daily, weekly, and monthly groups are exactly and exclusively chosen, we can perform hypothesis tests on the coefficient constraints implied by HAR model. If one sample survives after these tests, it would be in favor of HAR(1, 5, 22). Moreover, the time-varying (with rolling window analysis) accepting rates can be used as evidence to evaluate whether or not the classic specification assumed from investor behavior is appropriate and if so, precisely when.

The rest of the paper is arranged as follow. Section 2 gives the theoretical foundations of the Flexible HAR model, and the adaptive LASSO estimator and its statistical inference. Alternatives to be compared with our proposed model are presented in Section 3. Section 4 illustrates an empirical application with real high-frequency data of 10 individual stocks from the NYSE. Section 5 concludes.

2 Theoretical Foundations

2.1 HAR Model for Realized Volatility

Suppose the log-price X_t follows such standard continuous stochastic process

$$dX_t = \mu(t)dt + \sigma(t)dW_t, \tag{2.1}$$

where W_t is a standard Brownian motion, $\mu(t)$ is the trend which is a non random càdlàg finite variation process, and $\sigma(t)$ is the time-varying càdlàg volatility function independent of W_t.

The Integrated Volatility (IV) over one-day ($1d$) interval $[t-1d, t]$ is defined as

$$IV_t^{(d)} = \sqrt{\int_{t-1d}^{t} \sigma^2(u)du}. \tag{2.2}$$

The unobservable IV can be estimated by Realized Volatility (RV), which is calculated by the square root of the sum of squared log-returns over one day, namely

$$RV_t^{(d)} = \sqrt{\sum_{i=0}^{N-1} r_{t-i\cdot\Delta}^2}, \tag{2.3}$$

where N denotes the number of intraday observations, $\Delta = 1d/N$, $r_{t-i\cdot\Delta} = X_{t-i\cdot\Delta} - X_{t-i\cdot\Delta-\Delta}$. $RV_t^{(d)}$ converges to $IV_t^{(d)}$ in probability, as has been shown in **?**.

RV over longer time horizons (e.g. weekly and monthly with 5 and 22 trading days, respectively) are given as the average of daily RV over given periods

$$RV_t^{(w)} = \frac{1}{5}\left(RV_t^{(d)} + RV_{t-1d}^{(d)} + \ldots + RV_{t-4d}^{(d)}\right), \tag{2.4}$$

$$RV_t^{(m)} = \frac{1}{22}\left(RV_t^{(d)} + RV_{t-1d}^{(d)} + \ldots + RV_{t-21d}^{(d)}\right). \tag{2.5}$$

In addition, the Realized Kernel is a robust estimator of IV, even when returns are contaminated with noise; see **?** for more details. Note that the focus of this paper is modelling IV rather than estimating it. Therefore, we concentrate on only one accurate estimator of IV and the results should not be sensitive to any other choice.

The partial volatility at different time scales (monthly, weekly and daily) $\tilde{\sigma}_t^{(\cdot)}$ is assumed to follow a cascade structure of three additive equations with past realized volatility at the same time scale and expectation for the next period at a longer time scale.

$$\tilde{\sigma}_{t+1m}^{(m)} = \alpha^{(m)} + \rho^{(m)} RV_t^{(m)} + \tilde{\omega}_{t+1m}^{(m)}, \tag{2.6}$$

$$\tilde{\sigma}_{t+1w}^{(w)} = \alpha^{(w)} + \rho^{(w)} RV_t^{(w)} + \gamma^{(w)} \mathsf{E}_t\left[\tilde{\sigma}_{t+1m}^{(m)}\right] + \tilde{\omega}_{t+1w}^{(w)}, \tag{2.7}$$

$$\tilde{\sigma}_{t+1d}^{(d)} = \alpha^{(d)} + \rho^{(d)} RV_t^{(d)} + \gamma^{(d)} \mathsf{E}_t\left[\tilde{\sigma}_{t+1w}^{(w)}\right] + \tilde{\omega}_{t+1d}^{(d)}, \tag{2.8}$$

where $\tilde{\omega}_{t+1m}^{(m)}$, $\tilde{\omega}_{t+1w}^{(w)}$ and $\tilde{\omega}_{t+1d}^{(d)}$ are contemporaneously and serially independent zero-mean innovations.

Recursively substituting from (2.6) to (2.7) then to (2.8), and recalling that $\tilde{\sigma}_t^{(d)} = \sigma_t^{(d)}$

yields

$$\sigma_{t+1d}^{(d)} = \beta_0 + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \tilde{\omega}_{t+1d}^{(d)}. \tag{2.9}$$

with $\beta_0 = \alpha^{(d)} + \gamma^{(d)}\alpha^{(w)} + \gamma^{(d)}\gamma^{(w)}\alpha^{(m)}$, $\beta^{(d)} = \rho^{(d)}$, $\beta^{(w)} = \gamma^{(d)}\rho^{(w)}$, $\beta^{(m)} = \gamma^{(d)}\gamma^{(w)}\rho^{(m)}$.

Moreover, given that

$$\sigma_{t+1d}^{(d)} = RV_{t+1d}^{(d)} + \omega_{t+1d}^{(d)}, \tag{2.10}$$

where $\omega_t^{(d)}$ is the measurement error, and substituting (2.10) into (2.9) yields the HAR(1, 5, 22) model (by **?**)

$$RV_{t+1d}^{(d)} = \beta_0 + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \omega_{t+1d}, \tag{2.11}$$

where $\omega_{t+1d} = \tilde{\omega}_{t+1d}^{(d)} - \omega_{t+1d}^{(d)}$. Note that the HAR(1, 5, 22) model can also be rewritten as an AR(22) model

$$RV_{t+1d}^{(d)} = \theta_0 + \sum_{j=1}^{22} \theta_i RV_{t-(j-1)d}^{(d)} + \omega_{t+1d}, \tag{2.12}$$

with the constraints

$$\theta_j = \begin{cases} \beta^{(d)} + \frac{1}{5}\beta^{(w)} + \frac{1}{22}\beta^{(m)} & \text{for } j = 1; \\ \frac{1}{5}\beta^{(w)} + \frac{1}{22}\beta^{(m)} & \text{for } j = 2, \ldots, 5; \\ \frac{1}{22}\beta^{(m)} & \text{for } j = 6, \ldots, 22. \end{cases} \tag{2.13}$$

We extend (2.11) into a more general HAR$(1, \ldots, p)$ specification with p components in the model:

$$RV_{t+1d}^{(d)} = \beta_0 + \sum_{i=1}^{p} \beta_i \sum_{j=1}^{i} RV_{t-(j-1)d}^{(d)} + \omega_{t+1d}. \tag{2.14}$$

Obviously (2.14) can also be rewritten as a constrained AR(p) model.

2.2 Estimation of Flexible HAR

Under flexible HAR$(1, \dots, p)$ we do not the assume number of components to be included in the model and therefore use penalized regression (i.e. Least Absolute Shrinkage and Selection Operator, LASSO) to choose the active terms in the linear additive specification. The adaptive LASSO estimator is defined as

$$\widehat{\beta}_{AL} = \arg\min_{\beta} \left\{ \sum_{t=p}^{T} \left(RV_{t+1d}^{(d)} - \beta_0 - \sum_{i=1}^{p} \beta_i \sum_{j=1}^{i} RV_{t-(j-1)d}^{(d)} \right)^2 + \lambda \sum_{i=1}^{p} \lambda_i |\beta_i| \right\}, \tag{2.15}$$

where $\lambda \geq 0$ is the tuning parameter, which controls how strictly the penalization will be performed. The extreme case is $\lambda = 0$, which leads to the ordinary least squares (OLS) estimator. If $\lambda > 0$, all the truly non-zero coefficients will be penalized. Increasing λ causes fewer variables to be chosen. λ_i is the weight for each coefficient, which is data driven, e.g. the inverse of the absolute value of the corresponding OLS or ridge regression estimator. Ordinary LASSO introduced by **?** is a special case of the adaptive LASSO generalized by **?** with $\lambda_i = 1, \forall i = 1, \dots, p$.

Compared with **?** and **?**, who employ the LASSO on AR framework, here the penalized β in our model still keep the HAR structure and it is more convenient to conduct further tests on the structure.

2.3 Statistical Inference

Adaptive LASSO and its oracle properties were first introduced by **?** for cross-sectional data (i.i.d.). The estimators can identify the truly non-zero coefficients not only consistently but also asymptotically efficiently. **?** further derive the oracle properties of adaptive LASSO estimators for time series regression models:

- Consistency (variable selection):

$$\lim_{n\to\infty} \mathrm{P}\left(\widehat{\beta}_{AL} = \beta\right) = 1. \tag{2.16}$$

- Asymptotic normality for non-zero coefficient estimators $\widehat{\beta}^A_{AL}$:

$$\sqrt{n}\left(\widehat{\beta}^A_{AL} - \beta^A\right) + \widehat{b}^A_{AL} \xrightarrow{\mathcal{L}} \mathrm{N}\left(0, V^A\right), \tag{2.17}$$

  where the bias term $\widehat{b}^A_{AL} = \mathcal{O}\left(n^{-1/2}\right)$ and V^A is the covariance matrix. See Theorem 3.1 of **?**.

Moreover, they also investigate how to conduct statistical inference on the parameters which are not truly non-zero, i.e. how to test for false positives. Corollary 4.1 in their paper shows that for testing $H_{0,i} : \beta_i = 0$ versus $H_{1,i} : \beta_i \neq 0$, for $i \in \{1, \ldots, p\}$, the statistic $T_{\lambda,i} = \sqrt{n}|\widehat{\beta}_i|$ has the correct asymptotic size, where adaptive LASSO estimator $\widehat{\beta}_i$ depends on fixed λ

$$\lim_{n\to\infty} \sup_{0\leq\lambda<\infty} \mathrm{P}_{H_{0,i}}\left(T_{\lambda,i} > z_{i,1-\alpha}\right) \leq \alpha, \tag{2.18}$$

where $z_{i,1-\alpha}$ is the $1-\alpha$ quantile of the asymptotic distribution of the OLS estimator i.e. $\lambda = 0$. Based on these theoretical results, we can construct individual tests $H_0 : \beta_i = 0, \forall i = 1, \ldots, p$ by $T_{\lambda,i} = \sqrt{n}|\widehat{\beta}_i|$ and the standard normal distribution.

To test the validity of the lag structure by HAR(1, 5, 22), two kinds of null hypotheses can be tested, respectively. In particular, if the individual tests of $H_0 : \beta_i = 0, \forall i = 1, \ldots, 22$ can all be rejected, and in addition the joint test of $H_0 : \beta_{23} = \beta_{24} = \ldots = 0$ can be jointly accepted, then it should be a confirmation of the classic HAR specification.

To test whether all the coefficients beyond the 22nd are jointly significant we follow the same stepwise procedure as in **?**. For the multiple joint test with a large number of hypotheses, all the false hypotheses are desired to be rejected according to a sequence of threshold values given the significance level α:

- sort the p-values of the individual s statistics: $\widehat{p}_1 \leq \widehat{p}_2 \leq \ldots \leq \widehat{p}_s$
- if $\widehat{p}_1 \geq \alpha/s$, non-reject $H_{0,1}, \ldots, H_{0,s}$ and stop; otherwise reject $H_{0,1}$ and continue
- if $\widehat{p}_2 \geq \alpha/(s-1)$, non-reject $H_{0,2}, \ldots, H_{0,s}$ and stop; otherwise reject $H_{0,2}$ and continue

- ···

? suggest that such a stepwise multiple testing procedure could control the familywise error rate and capture the joint dependence structure of the test statistics. But as remarked on in **?**, the tests are very conservative, given that the individual tests are already conservative.

3 Alternative Models

To justify the performance of the proposed Flexible HAR model, we choose classical HAR(1, 5, 22), AR-AIC, AR-LASSO, HAR(a, b, c) with three non-zero coefficients, nonparametric HAR (HAR-NP), HARQ and HARQ-LASSO models as alternatives. They are all specified to model $RV^{(d)}_{t+1d}$ in different ways. Details of these alternatives are given as follow:

- **HAR(1, 5, 22)**: with classic lag structure (1, 5, 22), estimated by OLS, **?**

$$RV^{(d)}_{t+1d} = \beta_0 + \beta^{(d)} RV^{(d)}_t + \beta^{(w)} RV^{(w)}_t + \beta^{(m)} RV^{(m)}_t + \omega_{t+1d}. \tag{3.1}$$

- **AR-AIC**: select p by AIC, estimated by Yule-Walker equations

$$RV^{(d)}_{t+1d} = \theta_0 + \sum_{i=1}^{p} \theta_i RV^{(d)}_{t-(i-1)d} + \omega_{t+1d}. \tag{3.2}$$

- **AR-LASSO**: select λ by cross-validation

$$RV^{(d)}_{t+1d} = \theta_0 + \sum_{i=1}^{p} \theta_i RV^{(d)}_{t-(i-1)d} + \omega_{t+1d}, \tag{3.3}$$

$$\widehat{\theta} = \arg\min_{\theta} \left\{ \sum_{t=p}^{T} \left(RV^{(d)}_{t+1d} - \theta_0 - \sum_{i=1}^{p} \theta_i RV^{(d)}_{t-(j-1)d} \right)^2 + \lambda \sum_{i=1}^{p} |\theta_i| \right\}. \tag{3.4}$$

This was first employed by **?** and **?** as a flexible framework to determine the lag terms. Compared with our proposed Flexible HAR model, the HAR structure might be broken here after variable selection, since the constraints (2.13) on the AR coefficients cannot always hold in this case. Therefore we suggest that it would be more straightforward to perform hypothesis testing on HAR rather than AR terms and then compare with classical HAR(1, 5, 22).

- **HAR(a, b, c)**: select the smallest λ^* resulting in only three non-zero betas in the regularization path (shows how $\widehat{\beta}$ changes along with different λ values; one example can be found in Figure 4.2)

$$RV_{t+1d}^{(d)} = \beta_0 + \sum_{i=1}^{p} \beta_i \sum_{j=1}^{i} RV_{t-(j-1)d}^{(d)} + \omega_{t+1d}, \tag{3.5}$$

$$\widehat{\beta} = \arg\min_{\beta} \left\{ \sum_{t=p}^{T} \left(RV_{t+1d}^{(d)} - \beta_0 - \sum_{i=1}^{p} \beta_i \sum_{j=1}^{i} RV_{t-(j-1)d}^{(d)} \right)^2 + \lambda^* \sum_{i=1}^{p} |\beta_i| \right\}. \tag{3.6}$$

This alternative gives the fixed number of components to be included but not exactly which ones to be chosen. Furthermore, compared with **?**, who use all possible combinations of three numbers (within a maximum value), the LASSO method can provide a more efficient and data-driven approach.

- **HAR-NP**:

$$RV_{t+1d}^{(d)} = m_0 + m^{(d)}\left(RV_t^{(d)}\right) + m^{(w)}\left(RV_t^{(w)}\right) + m^{(m)}\left(RV_t^{(m)}\right) + \varepsilon_{t+1d}, \tag{3.7}$$

with $\mathsf{E}[\varepsilon_{t+1d}|\mathcal{F}_t] = 0$. m_0 is a constant, $m^{(\cdot)}(\cdot)$ are three smooth nonparametric link functions, which can be estimated by the Nadaraya-Watson smooth backfitting procedure introduced by **?**.

? propose an additive nonparametric extension on the HAR model and also the tests for linear parametric specifications. Their results show that the linearity assumption is widely rejected.

- **HARQ**:

Recent research by **?** argues that in practice data limitations put an upper bound on N and the resulting estimation error in RV might affect the coefficients in the HAR model. According to the asymptotic theory given in **?**,

$$\left(RV_t^{(d)} - IV_t^{(d)}\right) \xrightarrow{\mathcal{L}} \mathrm{N}\left(0, 2\Delta IQ_t^{(d)}\right), \text{ as } \Delta \to 0, \tag{3.8}$$

Integrated Quarticity (IQ), which reflects the asymptotic variance of the estimation error, should also be taken into account when modelling RV, namely the HARQ model. In parallel with IV and RV, Realized Quarticity (RQ) is also a consistent estimator of IQ.

$$IQ_t^{(d)} = \int_{t-1d}^{t} \sigma^4(u)du, \ RQ_t^{(d)} = \sum_{i=0}^{N-1} r_{t-i\cdot\Delta}^4. \tag{3.9}$$

$$\begin{aligned} RV_{t+1d}^{(d)} = \beta_0 &+ \left(\beta^{(d)} + \beta_Q^{(d)}\sqrt{RQ_t^{(d)}}\right) RV_t^{(d)} + \left(\beta^{(w)} + \beta_Q^{(w)}\sqrt{RQ_t^{(w)}}\right) RV_t^{(w)} \\ &+ \left(\beta^{(m)} + \beta_Q^{(m)}\sqrt{RQ_t^{(m)}}\right) RV_t^{(m)} + \omega_{t+1d}. \end{aligned} \tag{3.10}$$

RQ over one week and one month are defined in a similar way as (2.4) and (2.5). This model can be estimated by OLS and the parameters in brackets are also dynamic varying with the time series of RQ. They claim that their HARQ model can significantly improve the accuracy of forecasting compared with the HAR model.

- **HARQ-LASSO**:

In line with the proposed Flexible HAR model, the generalized HARQ model is given by

$$RV_{t+1d}^{(d)} = \beta_0 + \sum_{i=1}^{p}\left(\beta_i + \beta_{i,Q}\sqrt{\sum_{j=1}^{i} RQ_{t-(j-1)d}^{(d)}}\right)\sum_{j=1}^{i} RV_{t-(j-1)d}^{(d)} + \omega_{t+1d}. \tag{3.11}$$

We group all the regressors for each $i = 1, \ldots, p$ and employ the group LASSO to identify the active groups. The group LASSO estimate is defined as the solution to

$$\min_{\delta}\left\{\frac{1}{2}\sum_{t=p}^{T}\left(RV_{t+1d}^{(d)}-\sum_{i=1}^{p}\delta_i X_{i,t}\right)^2+\lambda\sum_{i=1}^{p}\|\delta_i\|_{K_i}\right\}, \tag{3.12}$$

where $K_1,\ldots,K_i$ are p.d. matrices.
$X_{i,t} \stackrel{\text{def}}{=} \left(\sum_{j=1}^{i} RV_{t-(j-1)d}^{(d)}, \sqrt{\sum_{j=1}^{i} RQ_{t-(j-1)d}^{(d)}}\sum_{j=1}^{i} RV_{t-(j-1)d}^{(d)}\right)$ includes the regressors in group i and $\delta_i \stackrel{\text{def}}{=} (\beta_i, \beta_{i,Q})^\top$. The group Least Angle Regression Selection Algorithm by **?** can be used to obtain the estimator. If one group i is chosen, then both of the two terms in this group will be active.

4 Empirical Application

4.1 Data

The empirical study is performed on millisecond trade data of 10 individual stocks from the New York Stock Exchange (NYSE): Boeing (BA), IBM, Johnson & Johnson (JNJ), Coca-Cola (KO), Walmart (WMT), Caterpillar (CAT), Walt Disney (DIS), Pfizer (PFE), UnitedHealth Group (UNH), and Exxon Mobil (XOM), for the period from September 10, 2003, to August 31, 2015. The data set is obtained from NYSE's Trades and Quotes (TAQ) database. First, we clean the raw high frequency data by following these steps, as suggested by **?**:

- Only keep the transactions between 9:30-16:00 in each trading day (when the exchange is open) and with non-zero price;
- Delete transactions with a correction indicator;
- If multiple transactions have the same time stamp, replace all these with the median price.

Next, for each trading day, calculate daily realized volatility by (2.3) for the returns in every 5 minutes. See Table 4.1 and Figure 4.1 for the descriptive statistics and time series plot (taking IBM as an example).

Index	mean$\cdot 10^{-2}$	std.$\cdot 10^{-2}$	min$\cdot 10^{-3}$	max	skewness	kurtosis	observations
BA	1.288	0.689	3.644	0.075	3.204	16.541	3013
IBM	1.023	0.615	3.145	0.102	4.549	35.388	3013
JNJ	0.818	0.443	2.398	0.068	4.440	33.995	3014
KO	0.902	0.474	2.311	0.071	4.164	29.759	3014
WMT	0.989	0.528	3.158	0.089	4.078	31.856	3013
CAT	1.474	0.881	3.984	0.127	3.414	19.876	3015
DIS	1.230	0.707	3.541	0.090	3.662	21.376	3013
PFE	1.179	0.594	3.604	0.111	3.977	35.521	3015
UNH	1.497	0.962	4.487	0.139	3.632	23.581	3015
XOM	1.119	0.680	3.181	0.133	5.292	54.917	3014

Table 4.1: Descriptive statistics of the realized volatility data of the 10 individual stocks under consideration

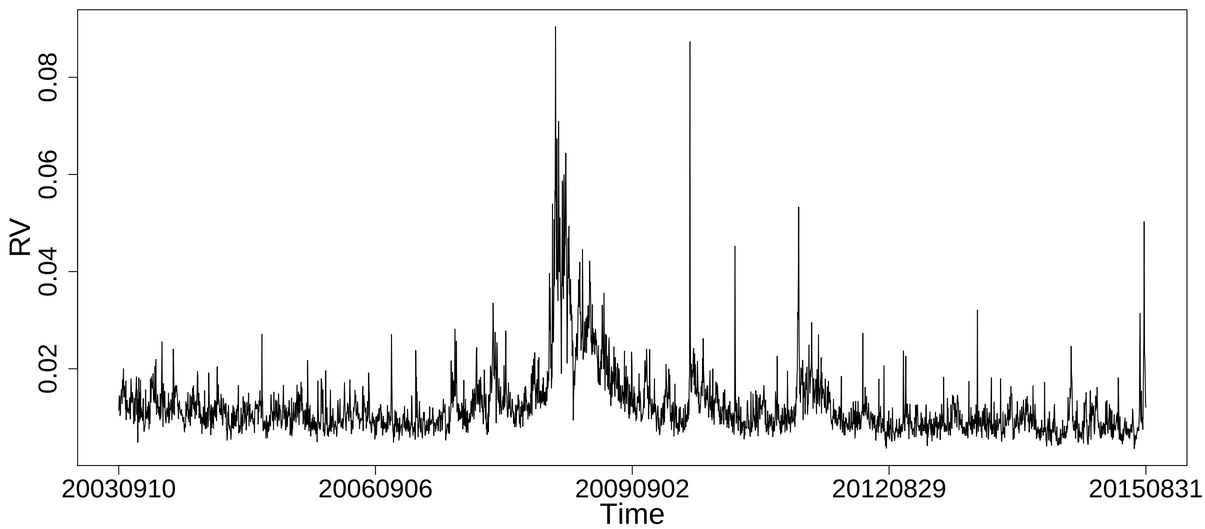

Figure 4.1: Time series of realized volatility for IBM, from September, 10 2003, to August 31, 2015, 3013 observations
f

4.2 Rolling Window Estimations and Testings

We set 1000 as the width of rolling windows in the empirical analysis. The maximum lag order is chosen as 50. The results are not sensitive to the arbitrary choice (also mentioned by ?). The weights for each coefficient in adaptive LASSO estimation are set as the inverse of the absolute value of the corresponding preliminary ridge regression estimator. The tuning parameter λ is chosen via cross-validation. In particular, we divide the whole sample $\{(x_t, y_t)\}_{t=1}^{T}$, into K (e.g. K=5) groups $G_1, \dots, G_K$ at random, hold out each group G_k, $k = 1, \dots, K$, at a time, construct an estimator on the remaining sample and predict the held out observations $\widehat{f}_{\lambda}^{-k}(x_t)$, for all $t \in G_k$. This procedure can be conducted for each value of tuning parameter λ. Hence, we can compute the mean squared error as a function of λ, namely the cross-validation error function

$$CV(\lambda) = \frac{1}{T} \sum_{k=1}^{K} \sum_{t \in G_k} \left\{ y_t - \widehat{f}_{\lambda}^{-k}(x_t) \right\}^2. \tag{4.1}$$

The optimal λ is chosen by minimizing $CV(\lambda)$

$$\widehat{\lambda}_{CV} = \arg\min_{\lambda} CV(\lambda). \tag{4.2}$$

Figure 4.2 illustrates the regularization path in Adaptive LASSO HAR estimation by (2.15) with the latest subsample (Aug. 29 2007 - Aug. 28 2015) for IBM. More estimated coefficients are penalized to be zero with higher tuning parameter λ. In this example, $\widehat{\lambda}_{CV} = 0.0167$ ($\log \widehat{\lambda}_{CV} = -4.0899$) and two terms with lag 21 and 30 are chosen.

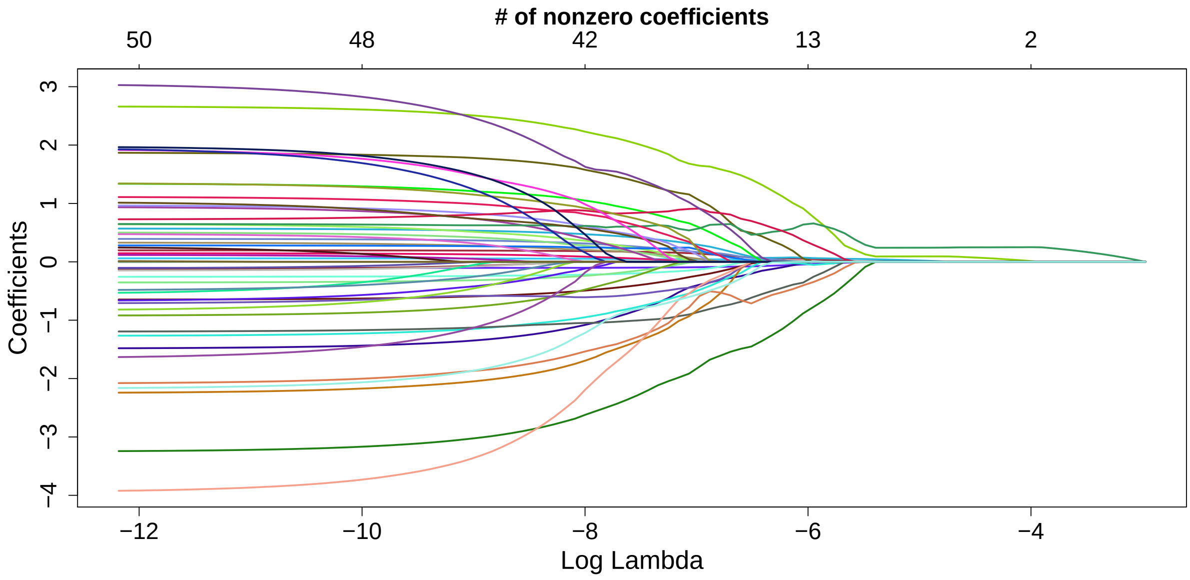

Figure 4.2: Estimated coefficients against the $\log \lambda$ sequence (subsample Aug. 29 2007 - Aug. 28 2015, IBM)

Consider all of the rolling window subsamples for IBM, in total the number of times each lag is non-zero (active) after LASSO estimation are counted; the percentages are shown in the red bars of Figure 4.3 and 4.4. Moreover, hypothesis testing on the coefficients, discussed in Section 2.3, is carried out. In greater detail, if the null hypothesis on each individual lag, $H_0 : \beta_i = 0, \forall i = 1, \ldots, 50$, cannot be rejected at the 95% confidence level in one subsample, although it is not shrunk to zero by penalization, its significance should be viewed as a false positive. The blue bars in the two plots report how many times each lag is significant both before and after individual tests. Similarly, one would also be counted in the grey bars in Figure 4.4 if it were active both before and after the joint tests on all lags beyond 22, $H_0 : \beta_{23} = \ldots = 0$.

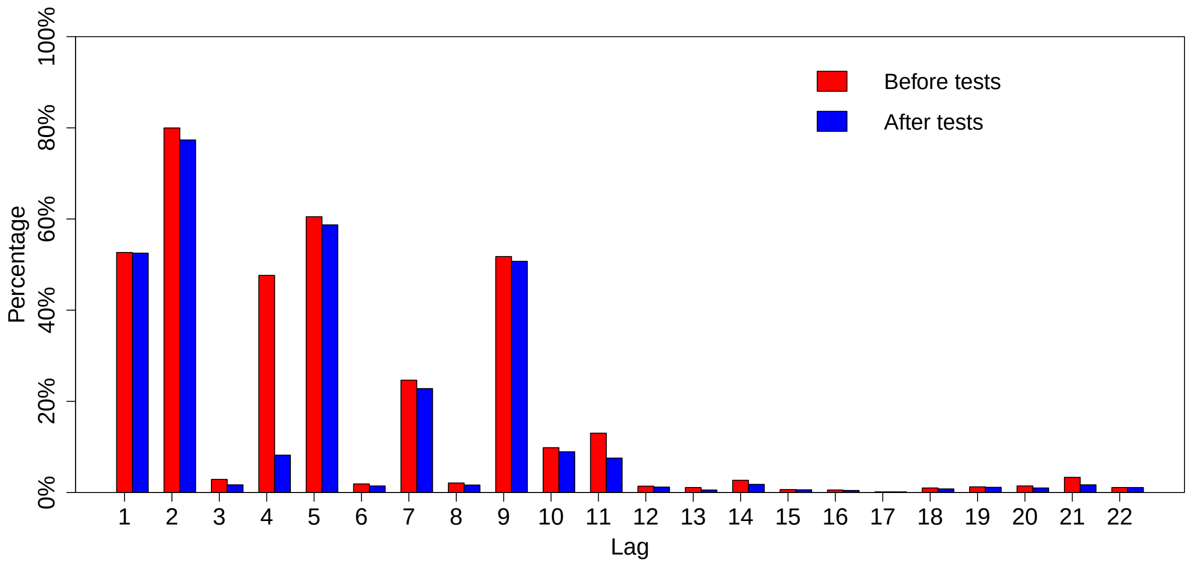

Figure 4.3: Percentage of time that each lag (1-22) is selected before (red) and after (blue) individual tests at 95% confidence level (results for IBM)

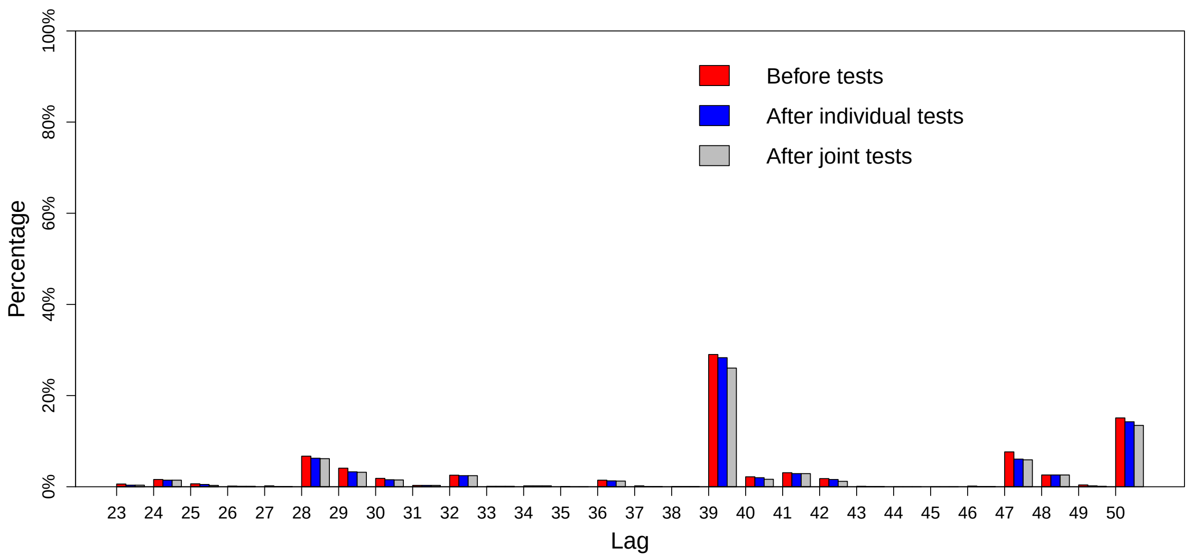

Figure 4.4: Percentage of time that each lag (23-50) is selected before (red) and after individual (blue) or joint (grey) tests at 95% confidence level (results for IBM)

Taking all stocks into consideration, Figure 4.5 shows the boxplot for the percentage of times that each lag (1-22) is selected by the adaptive LASSO in the flexible HAR$(1, \dots, p)$ model before and after the individual tests at a 95% confidence level . The difference before and after tests shows the false positive in LASSO estimation; e.g. it is quite obvious in lag 4.

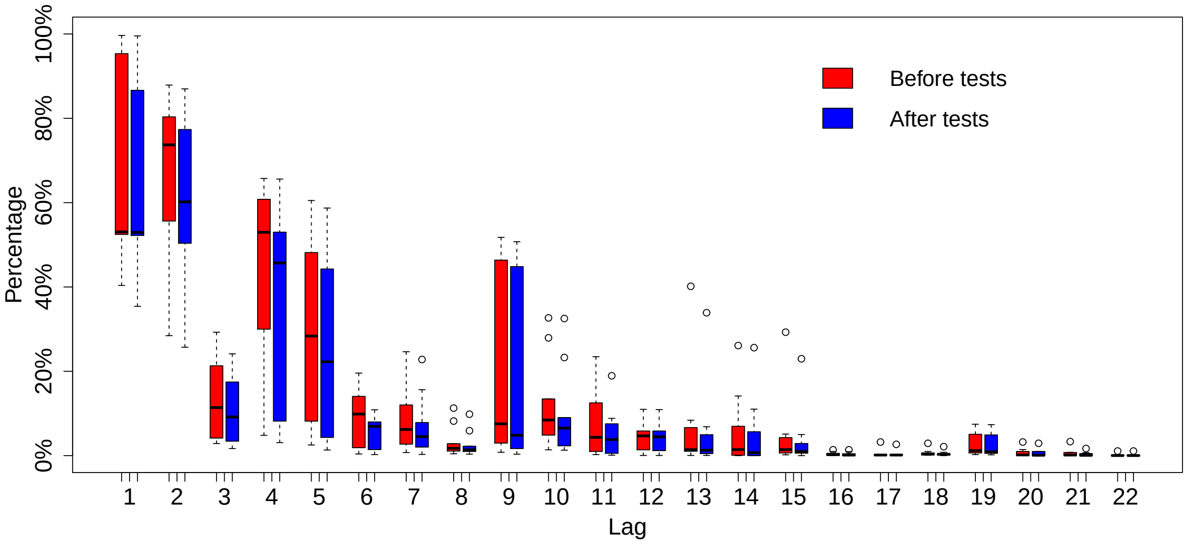

Figure 4.5: Percentage of time that each lag (1-22) is selected in Adaptive LASSO HAR before (red) and after (blue) individual tests at 95% confidence level (boxplot for all stocks)

The boxplot for the percentage of times that each lag (beyond 22) is selected by the adaptive LASSO in the flexible HAR$(1, \ldots, p)$ model before and after individual or joint tests at a 95% confidence level is displayed in Figure 4.6. As **?** mention, the joint test procedure is very conservative. Hence, we can see that the percentage of selection for large lag orders is much lower than under individual tests; in other words, more false positives would be detected by joint tests.

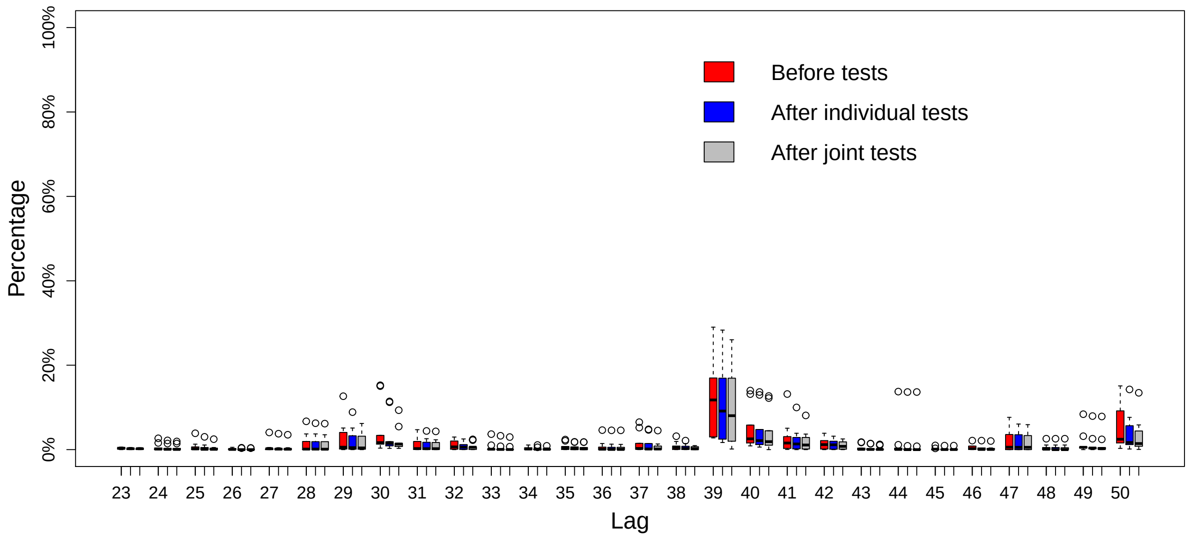

Figure 4.6: Percentage of time that each lag (23-50) is selected in Adaptive LASSO HAR before (red) and after individual (blue) or joint (grey) tests at 95% confidence level (boxplot for all stocks)

From all these plots, we find that most of the lags beyond 22 are not significant (except for lag 39 and the boundary 50). Concerning the choice of the maximum lag order, it indeed supports the HAR(1, 5, 22). But there is no uniform and strong evidence that the fixed lag structure can be exactly recovered by flexible models. It is especially questionable whether the monthly component under (1, 5, 22) should be included. In particular, there are also several small peaks in the plots, e.g. lag 9-11 (2 weeks), lag 19-20 (4 weeks), lag 29-31 (6 weeks) and lag 39-41 (8 weeks). This means the heterogeneous structure suggested by the classical HAR model indeed exists and can be recovered by flexible statistical models. However, the time scales for each component in the cascade could be longer or shorter (probably by 2 weeks rather than one month) and the classification of groups with different horizons in the market could somehow differ from their assumption. Even more importantly, the aim of our model is not to propose another fixed lag structure instead of (1, 5, 22). We prefer to choose a flexible specification completely driven by the data. Whether such a flexible HAR model can really outperform the classical HAR is further discussed below.

4.3 Estimation and Forecasting Accuracy

Here, we compare our flexible model with classic alternative models that were introduced in Section 3, in terms of the in-sample fitting and out-of-sample forecasting performance. We use Root Mean Square Error (RMSE) as the performance measure, which is defined as

$$RMSE = \sqrt{T^{-1}\sum_{t=1}^{T}(\widehat{RV}_t^{(d)} - RV_t^{(d)})^2}. \tag{4.3}$$

In particular, for in-sample fitting, we calculate the average RMSE (averaged over all rolling window subsamples for each individual stock) by

$$RMSE_{IS} = M^{-1}\sum_{m=1}^{M}\sqrt{(N-p)^{-1}\sum_{t=p+m}^{N+m-1}(\widehat{RV}_t^{(d)} - RV_t^{(d)})^2}. \tag{4.4}$$

For one step ahead out-of-sample forecasting, we compute the RMSE as

$$RMSE_{OS} = \sqrt{M^{-1}\sum_{m=1}^{M}(\widehat{RV}_{N+m}^{(d)} - RV_{N+m}^{(d)})^2}, \tag{4.5}$$

where p is the maximum lag, N is the width of each rolling window subsample and M is the number of rolling windows.

The boxplots for the comparison results among all stocks are illustrated in Figure 4.7 and 4.8.

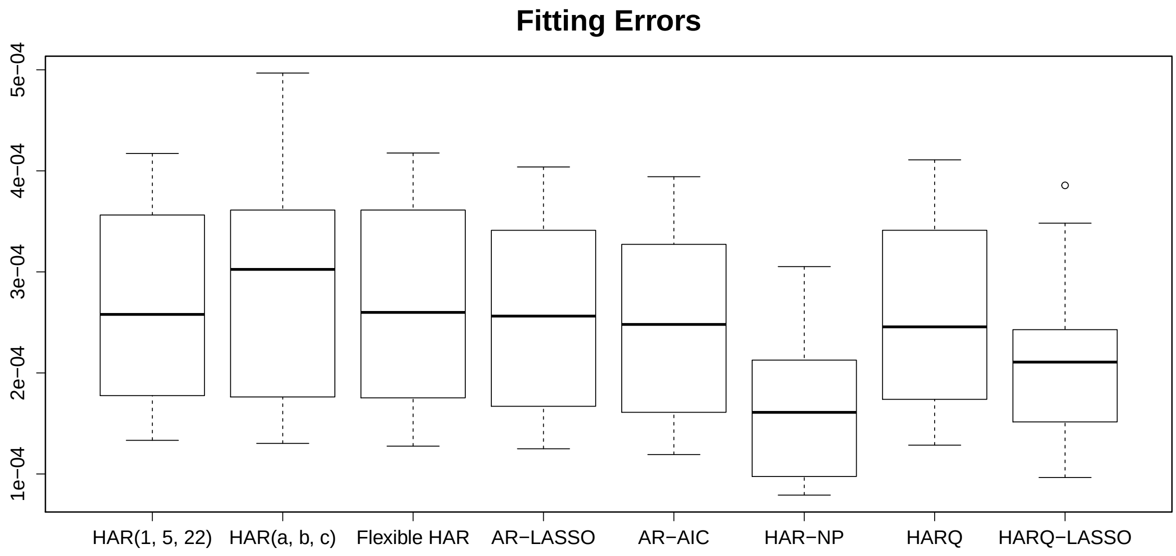

Figure 4.7: In-sample fitting errors under each model

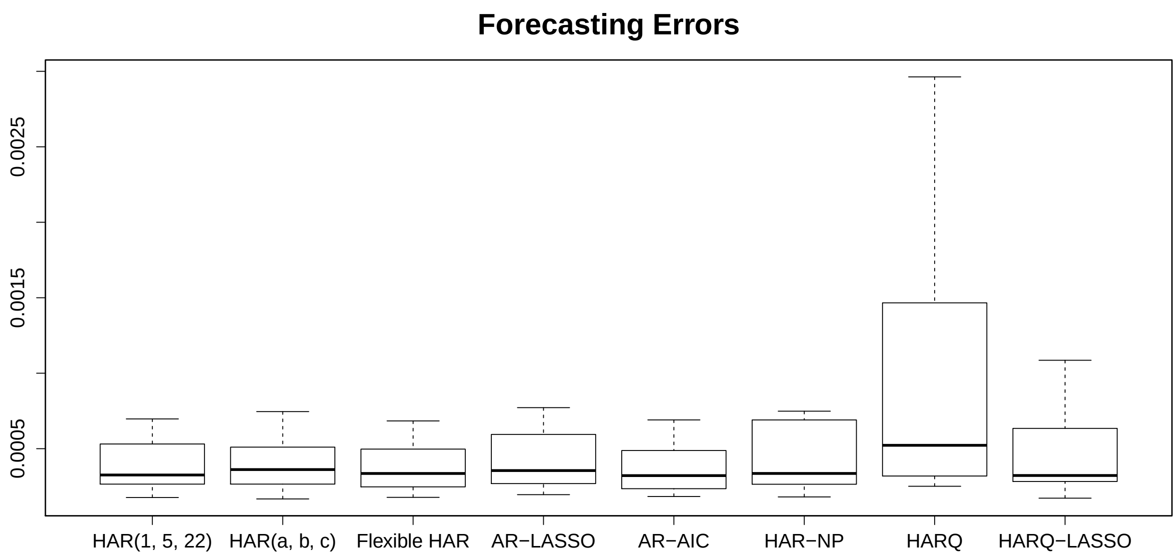

Figure 4.8: Out-of-sample forecasting errors under each model

Concerning in-sample fitting, nonparametric estimation performs the best. The HARQ model and all flexible lag structure specification are better than the fixed HAR (1, 5, 22) and HAR(a, b, c) models. However, in terms of the RMSE for out-of-sample data, the HARQ model is the worst one in forecasting. In our sample, the HARQ model does not work as well as **?** claimed in their paper. The main reason might be that we did not apply any "filter" for the ourliers in forecasting as the authors mentioned in footnote 15. This may be unfair to other models since the results from other models are quite robust even considering the original whole sample. Nonparametric and flexible HARQ models also show bad performance in forecasting. For a clear comparison, in Table 4.2 we also report the ratios of the errors in all the alternatives relative to our benchmark (Flexible HAR model), averaged over all individual stocks.

	HAR(1, 5, 22)	HAR(a, b, c)	Flexible HAR	AR-LASSO
$RMSE_{IS}$	1.003	1.099	1.000	0.976
$RMSE_{OS}$	1.020	1.046	1.000	1.144

	AR-AIC	HAR-NP	HARQ	HARQ-LASSO
$RMSE_{IS}$	0.934	0.652	0.968	0.819
$RMSE_{OS}$	0.990	1.175	2.379	1.156

Table 4.2: In-sample fitting and out-of-sample forecasting errors (ratios relative to proposed Flexible HAR model, averaged over all individual stocks)

Only AR-AIC is the best one for both in- and out-of-sample. Our proposed model performs the best in forecasting among all models under the HAR framework. In particular, compared with classic HAR(1, 5, 22), which is thought to be unbeatable in previous studies, our flexible model improves on it slightly. Our finding differs from the work by **?**, who use LASSO on AR framework as in (3.3). Note that without the coefficient restrictions as in (2.13), it is unlikely that the model after LASSOing (3.3) can be converted back to the HAR model. In other words, the penalized β in their model can no longer keep the HAR structure, whereas our proposed model does not have this problem.

To evaluate the volatility forecast performance formally, we also employ the test for superior predictive ability - the Hansen Test by **?**. The null hypothesis is given by, $H_0 : \mathsf{E}(L_{0,t} - L_{k,t}) \leq 0$, where $L_{\cdot,t}$ is the loss function, e.g. squared error loss. The test is implemented by bootstrap when choosing the critical values. **?** classified three types of tests, leading to different bootstrap distributions. Consequently, liberal, consistent and conservative tests give the lower bounds, consistent estimators and upper bounds for the true p values, respectively. We set the proposed flexible HAR model as the benchmark and test all the alternatives jointly. The results averaged over all stocks are shown in Table 4.3 (number of bootstrapped samples is 10,000).

	SPA_l	SPA_c	SPA_u
p values	0.486	0.760	0.857

Table 4.3: Superior Predictive Ability (SPA) tests results (averaged over all stocks)

To be more specific, 99.67% of the p values are higher than 5% in our sample, except for the lower bound for DIS. This implies that we cannot reject the null hypothesis significantly. Therefore, we can conclude that the flexible HAR model cannot be significantly outperformed by any of the competitors.

Additionally, the Model Confidence Set (introduced by **?**) is constructed through a sequence of significance tests. In each step, if the hypothesis of Equal Predictive Ability (EPA) is rejected (under significace level α), one is to remove the worst one found to be significantly inferior to any others, until EPA is accepted. As a result, the set of superior models {HAR(1, 5, 22), HAR(a, b, c), Flexible HAR, AR-AIC, HAR-NP} is identified with p value = 0.3784 by a bootstrap procedure of 10,000 resamples. During the sequence of testing, HARQ, LASSO-HARQ, AR-LASSO are eliminated in turn under a 5% significance level.

4.4 Further Validation of the HAR Structure

In addition, to test the validity of the classic HAR lag structure further, we group the lags in AR(50) as $\{1\}, \{2-5\}, \{6-22\}, \{23-50\}$ (probably smaller clusters beyond 22), and employ group LASSO to identify the active AR lags. Similarly to (3.12) the group LASSO estimate is defined as the solution to

$$\min_{\theta} \left\{ \frac{1}{2} \sum_{t=p}^{T} \left(RV_{t+1d}^{(d)} - \sum_{j=1}^{J} \theta_j X_{j,t} \right)^2 + \lambda \sum_{j=1}^{J} \|\theta_j\|_{K_j} \right\}, \tag{4.6}$$

where J is the number of groups. In this case, $J = 4$, $X_{j,t}$ includes the lag AR terms in group j and θ_j contains all the related coefficients.

Group LASSO estimation implies that if one group is active, then all the variables in it will be active. Thus, if the daily, weekly, and monthly groups are exactly and exclusively chosen, then we can perform hypothesis tests on the coefficient constraints implied by the HAR model as in (2.13). If one rolling window subsample survives after these tests, this would favor HAR(1, 5, 22). In our results, on average only 0.73% of the subsamples can survive such test procedures.

On average, 13.70% of the rolling window subsamples do not have significant lags after 22. Furthermore, only 9.09% of the subsamples have significant lags in the monthly group. Finally, 0.73% of the subsamples can still survive the coefficient constraint test. Therefore, we can conclude that the main reason for rejection might be that there are still some important lags beyond one month and the significance of the monthly group under HAR (1, 5, 22) is not certain. In that case, the minor reason is the equality constraints on the coefficients, i.e. taking a sample average when calculating the realized volatility over longer time horizons is suspect. Our proposed model can give a flexible specification when arranging the terms but still keep the HAR structure on coefficients. The reasoning behind our model is confirmed by these results.

Moreover, the time-varying (with rolling window analysis) accepting rates can be used as evidence to evaluate whether the classic specification assumed from investor behavior is appropriate and if so, precisely when. For each rolling window subsample, the accepting rate can be calculated as

$$ratio_t = \frac{\#\text{ of stocks that survive the test}}{\text{total }\#\text{ of stocks (10)}}$$

The time-varying accepting rates (averaged over all stocks) is shown in Figure 4.9.

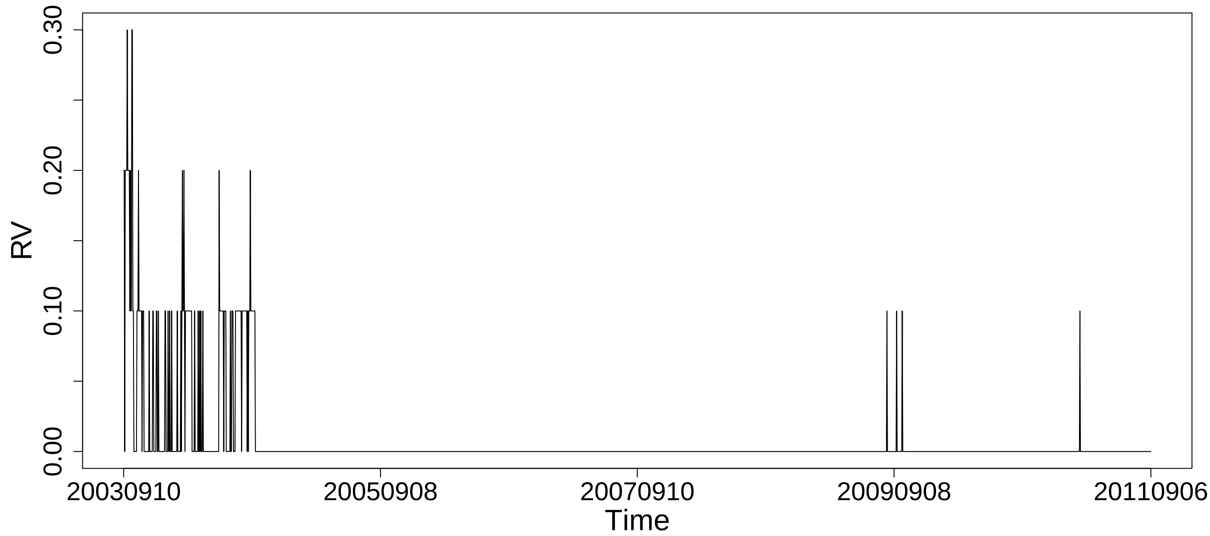

Figure 4.9: Time-varying accepting rates in the tests (averaged over all stocks)

The results of time-varying analysis show that relatively high accepting rates only occur in a short period at the beginning of our sample (2003-2004). The accepting rates are almost zero during a crisis. This means that when the market environment is not stable, the structure of (1, 5, 22) does not hold very well.

5 Conclusions

In this paper, we propose a more generalized and flexible HAR model for realized volatility dynamics. We employ the adaptive LASSO variable selection method and its statistical inference theory to choose the active components that need to be included in the HAR framework and to see whether the implied lag structure (1, 5, 22) from an economic point of view can be recovered by statistical models.

We use the daily realized volatility data for 10 individual stocks from 2003 to 2015 as the data set in the empirical analysis. The adaptive LASSO estimation and the subsequent hypotheses testing results for all rolling window subsamples show that there is no uniform and strong evidence that the lag structure (1, 5, 22) can be exactly recovered by flexible models. In particular, it is questionable whether the monthly component should be included. In addition, there are some small peaks every two weeks. It seems that the heterogeneous structure suggested by the HAR model does indeed exist but the time scales for each component in the cascade could be different from the classical simple assumption.

Furthermore, we compare our flexible model with some other alternatives in terms of in-sample fitting and out-of-sample forecasting performances. Based on the RMSE for out-of-sample data, our flexible model is not significantly outperformed by any of the alternatives. This conclusion is also supported by superior predictive ability tests.

In addition, we employ a group LASSO to identify the active AR lags and do some related tests to check the validity of the classic HAR lag structure. On average, only 0.73% of the subsamples survive after the whole testing procedures. With some further analysis, we conclude that the main reason for rejection might be the arrangement of groups, whereas the minor reason is the equality constraints on AR coefficients. Flexible arrangement of groups while still keeping the HAR frame is exactly what our proposed model strives to specify. Finally, the time-varying accepting rates show that relatively high accepting rates only occur in a short period at the beginning of our sample (2003-2004). When the market environment is not stable, the structure of (1, 5, 22) does not hold very well.