AIMS *Quantitative Finance and Economics*

http://www.aimspress.com/journal/QFE

QFE, 1(4): 363–387
DOI:10.3934/QFE.2017.4.363
Received date: 7 September 2017
Accepted date: 19 October 2017
Published date: 13 December 2017

Research article

Testing the Lag Structure of Assets' Realized Volatility Dynamics

Francesco Audrino*, **Lorenzo Camponovo and Constantin Roth**

Department of Economics, University of St. Gallen, Bodanstrasse 6, 9000 St.Gallen, Switzerland

* **Correspondence:** francesco.audrino@unisg.ch; Tel: +41-71-224-2431; Fax: +41-71-224-2894.

Abstract: A (conservative) test is applied to investigate the optimal lag structure for modeling realized volatility dynamics. The testing procedure relies on the recent theoretical results that show the ability of the adaptive least absolute shrinkage and selection operator (adaptive lasso) to combine efficient parameter estimation, variable selection, and valid inference for time series processes. In an application to several constituents of the S&P 500 index it is shown that (i) the optimal significant lag structure is time-varying and subject to drastic regime shifts that seem to happen across assets simultaneously; (ii) in many cases the relevant information for prediction is included in the first 22 lags, corroborating previous results concerning the accuracy and the difficulty of outperforming out-of-sample the heterogeneous autoregressive (HAR) model; and (iii) some common features of the optimal lag structure can be identified across assets belonging to the same market segment or showing a similar beta with respect to the market index.

Keywords: realized volatility; adaptive lasso; HAR model; test for false positives; lag structure

1. Introduction

For many years the analysis of the time-varying behavior of financial assets has been the goal of extensive research. While it seems of course tempting to be able to forecast future price movements and returns, there is consensus that a monetarily profitable way of forecasting returns does not exist; see, e.g., Goetzmann and Jorion (1993); Nelson and Kim (1993); Kothari and Shanken (1997); Stambaugh (1999); Torous et al. (2004); Campbell and Yogo (2006); Ang and Bekaert (2007), among others. On the other hand, however, it is a well known fact that the volatility of an asset's return series is very persistent and can be captured relatively well by several model types. The ability to estimate, at least approximately, the volatility of an asset's returns is interesting for a variety of practical reasons, most notably portfolio management and risk management. In particular, asset managers like pension funds or hedge funds usually profit from the ability to estimate volatility by being able to better adjust the risk-return tradeoff in their portfolios. Moreover, any institution holding financial assets, whether for

speculation or hedging, is capable of better calculating the risks that are introduced to their balance sheets by specific assets if volatility estimates are available. An important practical application of this is, for example, the Value-at-risk, which is crucially dependent on accurate volatility estimates and forms a major part of the rules imposed by financial regulators on individual institutions.

One of the earliest and most influential of these models is the well known GARCH model introduced by Bollerslev (1986). Although countless extensions and adaptations have been suggested, it is very difficult to outperform the classic GARCH (1,1) model in terms of forecasting quality (Hansen and Lunde, 2005).

In the last years, the focus of research on volatility has been extended beyond modeling volatility in general stochastic volatility (SV) or GARCH settings to the idea of computing accurate estimates of volatility using high-frequency data, introducing the notion of *realized volatility*; for more details, we refer to the seminal works of Andersen et al. (2001, 2003) or Barndorff-Nielsen and Shephard (2002), and for a more recent overview of the literature regarding realized volatility and its estimation to McAleer and Medeiros (2008). In order to capture long memory, one of the main stylized facts shown by assets' realized volatility time series, the time-varying dynamics of realized volatility are often modeled as an ARFIMA process; see, for example, Andersen et al. (2001). As an alternative, the most commonly used model for realized volatility is the heterogeneous autoregressive (HAR) model introduced by Corsi (2009). Although not formally belonging to the class of long memory models, the HAR model is able to closely mimic the behavior shown by realized volatility time series. Nowadays, the HAR model enjoys great popularity because of its very good out-of-sample forecasting performance and at the same time its computational simplicity. Similarly to the experience with the classic GARCH (1,1) model, the standard HAR model has been extended along different directions in an attempt to improve its estimation and forecasting accuracy; see Corsi et al. (2012) for a recent survey.

The aim of this paper is to shed some light about the (possibly time-varying) lag structure driving realized volatility dynamics. Moreover, making use of the fact that the HAR model can be viewed as a (restricted) autoregressive (AR) process of daily realized volatility, we also analyze whether the lag structure and the maximal order lag implied by the HAR model, which is derived based on theoretical assumptions about the investor behavior, can be recovered purely by statistical means. In order to do so, we will use the adaptive lasso introduced by Zou (2006), a refinement of the standard lasso technique due to Tibshirani (1996), as the estimation procedure and apply it to the set of past lags in an unrestricted AR model of a sufficiently large order. Additionally, we also see whether the lags that are significantly selected are in line with the HAR assumptions.*

For the analysis, we will make use of the recently developed theory by Audrino and Camponovo (2015) who derived a conservative testing procedure for the coefficients estimated by the adaptive lasso. As shown in Leeb and Pötscher (2006, 2008) and Pötscher and Schneider (2009) , among others, it is impossible to estimate the distribution of lasso estimators in the uniform sense (uniformity with respect to the underlying parameter of interest in a shrinking neighborhood of the origin). The conservative testing procedure introduced in Audrino and Camponovo (2015) aims to partially mitigate

*If the HAR model is the correct data generating process, applying the lasso we should expect to select lags up to order 22 in the active set, whereas the coefficients of the lags beyond 22 should be set to zero. Note that this is clearly different from testing the full HAR specification, as in this case the imposition of the equality restrictions across coefficients belonging to the same frequency (weekly, monthly) is necessary as well. This goal can be attained in the lasso context by considering, for example, the cluster group lasso (Bühlmann et al., 2013).

this problem. In particular, Audrino and Camponovo (2015) show that under appropriate conditions, their approach ensures uniform valid inference with respect to the selection of the adaptive lasso tuning parameter. While Zou (2006) was able to prove how to test hypotheses on the active set of parameters (the coefficients that are truly non-zero in the population), Audrino and Camponovo (2015) extended the theory to hypothesis testing on the passive set of variables (the coefficients that are truly identical to zero in the population), as well. This newly developed theory now makes it possible to test hypotheses regarding the coefficients and therefore to analyze which lags are significantly different from zero. For a comprehensive review of the literature regarding inference of the adaptive lasso, we refer to Audrino and Camponovo (2015).

Surprisingly, thus far there has been little research on the topic of analyzing the validity of the assumptions of the HAR model. While Craioveanu and Hillebr (2012) have shown that it is only of little importance what investment horizons are assumed for the different investor groups, that is which aggregation frequencies are chosen in a HAR-like structure, Wang et al. (2013) show that an ARFIMA model subject to structural breaks in the mean and the memory parameters of the process can be approximated well by an AR model. They interpret this result as an indirect econometric explanation for the empirical success of the HAR model. Finally, Hwang and Shin (2014) develop an infinite-order extension of the HAR, called HAR (∞), and show that in estimating it using a finite-order model HAR, the prediction error is mainly due to the estimation of the HAR coefficients rather than to errors made in approximating the HAR (∞) process by the misspecified finite-order HAR model, thus providing a theoretical justification for the wide use of the standard HAR model.

The present paper goes in the same direction as the work by Audrino and Knaus (2016) who also analyze the optimal lag structure and maximal order lag of realized volatility dynamics in an unrestricted autoregressive setting. They adopt a similar approach and choose the lasso estimation procedure to recover the lags that seem most informative for predicting future realized volatility. In an out-of-sample analysis, the authors show that it is difficult to outperform the benchmark HAR model when applying a theoretically sound variable selection procedure in an unrestricted AR setting. This paper differs from Audrino and Knaus (2016) along several dimensions and, moreover, addresses the various open points mentioned in their concluding remarks section. First, Audrino and Knaus (2016) only analyze which lags are selected from the lasso but do not test formally for the presence of false positives. We extend the analysis by considering the adaptive lasso, a refinement of the lasso that should itself help reduce the number of false positives, and additionally perform a test on the significance of the coefficients in a statistically sound way. As we will show, the presence of false positives is pervasive in the analysis of the lag structure of realized volatility and, thus, must be taken seriously into account. Otherwise, we run the risk of drawing the wrong conclusions.

Second, we extend the data set cross-sectionally by considering more individual stocks belonging to the S&P 500. This allows us to better investigate the evolution of the structure of the selected lags over time. In contrast to Audrino and Knaus (2016) we can (i) verify whether there are common structural breaks affecting simultaneously the dynamics of (almost) all assets and (ii) determine the factors, if any exist, that are able to explain the cross-sectional differences in how structural breaks influence the lag structure. Our results show that there are two main structural breaks in the period under investigation ranging from January 2001 to November 2010 that divide it into three different subperiods characterized by different market and realized volatility conditions: The first break corresponds to the end of the US stock market downturn in 2002 following the burst of the dot-com

bubble, whereas the second break takes place with the onset of the financial crisis in 2008, located around the bankruptcy of Lehmann Brothers. Moreover, we show that the market sector and the beta of the individual assets are important factors in explaining the behavior cross-sectionally of the realized volatility lag structures: in fact, assets that belong to the same market sector or possess a similar beta with respect to the market index show very similar structures.

The content of this paper can be summarized as follows: Section 2 presents the theoretical foundations of both the HAR model and the adaptive lasso estimator and introduces the hypotheses of interest and the testing procedure. Section 3 illustrates the empirical results of the analysis by separately presenting them before and after testing. Section 4 concludes.

2. Theoretical Considerations

This section will lay out the theoretical motivation of the empirical tests that follow. The idea is to estimate the optimal lag structure driving realized volatility dynamics and to test the validity of the lag structure and maximal order lag implied by the HAR model by employing the adaptive lasso as model selection device. The estimation results of the adaptive lasso will further be analyzed regarding statistical significance both individually and jointly. We will therefore first introduce the HAR model, then the estimation procedure via the adaptive lasso and the conditions under which inference can be conducted. Finally, the concept of multiple hypothesis testing will be briefly revisited.

2.1. The HAR model

In the past years, there has been a shift in the focus of the literature on modeling and forecasting volatility. For many years after the introduction of the GARCH model by Bollerslev (1986), the literature aimed at modeling the unobservable volatility of an asset. However, in recent years, the concept of *realized volatility* has gained popularity: in fact, by estimating the unobservable second moment more accurately using high-frequency data, volatility has been made somehow "observable". Let us briefly review the concept of realized volatility.

We assume that log-prices of an asset follow the stochastic process

$$dp(t) = \mu(t)dt + \sigma(t)dW(t),$$

where $W(t)$ is a standard Brownian motion and $\mu(t)$, $p(t)$ and $\sigma(t)$ satisfy the usual regularity conditions. Under this assumption, realized volatility is defined as the sample analogue to the quadratic variation $\sqrt{\int_{t-1d}^{t} \sigma^2(s)ds}$, where d corresponds to the time interval of interest, usually one trading day.

As has been shown in previous research on quadratic variation by Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2002), this theoretical quantity can be estimated consistently by means of the sum of squared log-returns, assuming the absence of microstructure noise and jumps. With appropriately chosen Δ and M, which reflect the length of the sampled time interval and the sampling frequency, one can therefore define realized volatility as

$$RV_t^{(d)} = \sqrt{\sum_{j=0}^{M-1} r_{t-j\cdot\Delta}^2},$$

where r_t are the asset's log-returns.

When it comes to modeling realized volatility dynamics over time, one of the most popularly used models for its simplicity and at the same time very high estimation and prediction accuracy, is the *heterogeneous autoregressive model*, henceforth the HAR model, introduced by Corsi (2009). The central idea of the HAR model is to predict future realized volatility dependent on past realized volatility in a linear setting.

The HAR model is stated under the key assumption that the market participants can be separated into three heterogeneous groups of traders according to their trading horizons, i.e. daily, weekly or monthly rebalancing. It further assumes that each of these groups create their own component of latent volatility and take into account their expectation of future volatility at a lower frequency. This cascade model of volatility can finally be written as the famous HAR equation. Furthermore, it is common to substitute realized volatility at different frequencies by the respective logarithms. This is done mainly for two reasons: First, the distribution of the logarithm of realized volatility more closely resembles the normal distribution than realized volatility in its level form. Second, by modeling the logarithm there is no need to impose restrictions on the parameters so that resulting estimates and forecasts are positive. After this transformation, we get:

$$\log RV_{t+1}^{(d)} = c + \beta^{(d)} \log RV_t^{(d)} + \beta^{(w)} \log RV_t^{(w)} + \beta^{(m)} \log RV_t^{(m)} + \epsilon_{t+1} \tag{1}$$

Using the fact that the logarithms of weekly and monthly realized volatility, $\log RV_t^{(w)}$ and $\log RV_t^{(m)}$, are defined to be the averages of the logarithms of daily volatility over 5 and 22 days (the number of trading days in a week/month), respectively, this model can be rewritten as a constrained linear autoregressive (AR) process:

$$\log RV_{t+1}^{(d)} = \beta_0^{HAR} + \sum_{i=1}^{22} \beta_i^{HAR} \log RV_{t-i+1}^{(d)} + \epsilon_{t+1} \tag{2}$$

with the restrictions being

$$\beta_i^{HAR} = \begin{cases} \beta^{(d)} + \frac{1}{5}\beta^{(w)} + \frac{1}{22}\beta^{(m)} & \text{for } i = 1; \\ \frac{1}{5}\beta^{(w)} + \frac{1}{22}\beta^{(m)} & \text{for } i = 2, ..., 5; \\ \frac{1}{22}\beta^{(m)} & \text{for } i = 6, ..., 22. \end{cases} \tag{3}$$

In inspecting the structural AR (22) specification implied by the HAR model, it is obvious that if the HAR model were the correct model for realized volatility, one would expect to observe in empirical data that indeed exactly the first 22 lags are relevant for forecasting tomorrow's realized volatility.[†]

It is now the goal to apply the adaptive lasso as a model selection device to the time series of realized volatility in an unrestricted AR model of a sufficiently large order. If the HAR model were the correct model, we would expect the adaptive lasso to select exactly the first 22 lags and set the remaining coefficients to zero. While this question was previously of interest for Audrino and Knaus (2016), our procedure differs in two ways: First, we will use the adaptive lasso, as opposed to the standard lasso, which was used by Audrino and Knaus (2016) in the main analysis. Second, we will additionally apply the newly derived theory by Audrino and Camponovo (2015) in order to check which of the lags that

[†]From the original assumptions it can be derived that all $\beta^{(\cdot)}$ are strictly positive, which rules out the scenario of any β_i^{HAR} being 0. This has been confirmed empirically.

have been selected by the adaptive lasso are truly significant. In this way we will get rid of, or at least reduce the impact of, false positives (that is, variables that are selected to be active but are in reality inactive) that affect lasso-type approaches and can possibly lead to wrong conclusions.

2.2. *The adaptive lasso*

The adaptive lasso, as proposed by Zou (2006), is a generalized refinement of the lasso that was originally introduced by Tibshirani (1996). The lasso, an acronym for least absolute shrinkage and selection operator, can be viewed as a constrained least squares regression problem. Since the model to be tested is linearly autoregressive in the regressors, i.e.

$$\log RV_{t+1}^{(d)} = \beta_0 + \sum_{i=1}^{p} \beta_i \log RV_{t-i+1}^{(d)} + \epsilon_{t+1}, \tag{4}$$

where the ϵ_i are innovation variables with mean zero, we can consequently define the adaptive lasso estimator as

$$\hat{\beta}_{AL} = \arg\min_{\beta} \left\{ \sum_{i=p}^{n} \left(\log RV_{i+1} - \beta_0 - \sum_{j=1}^{p} \beta_j \log RV_{i-j+1}^{(d)} \right)^2 + \lambda \sum_{j=1}^{p} \lambda_j |\beta_j| \right\} \tag{5}$$

with $\lambda \geq 0$ being a tuning parameter and λ_j being individual weights for each of the coefficients. The special case of the classic lasso by Tibshirani (1996) corresponds to the setting where $\lambda_j = 1$ for $j = 1, ..., p$.

From this objective function it is obvious that setting $\lambda = 0$ will cause the lasso estimator to coincide with the ordinary least squares estimator, $\hat{\beta}_{OLS}$. Usually, this estimator will not set any of the coefficients equal to zero. However, by choosing a strictly positive λ, all coefficients that are not equal to zero will be penalized. Consequently, increasing λ leads to more and more coefficients that will be set exactly to zero, therefore performing progressively stricter variable selection. Thus, assuming sparsity, we expect the adaptive lasso to select only the active set of regressors, i.e. the ones which are truly non-zero.

For the classical cross-sectional setting, assuming the innovation variables to be independent and identically distributed (iid), Zou (2006) showed that the adaptive lasso fulfills the so-called *oracle properties*, which means that it asymptotically both identifies the non-zero coefficients and estimates the coefficients of the truly active variables consistently and efficiently. This result is dependent on the fact that the individual weights λ_j will be chosen data-dependent. Possible choices for the weights are, for example, the inverse of the absolute value of the coefficients estimated by either OLS (in the case that OLS can be performed) or the lasso (in the high-dimensional case where one has to deal with more regressors than observations).

However, for the specific question we are analyzing, the assumption of iid innovations is unrealistic. When it comes to linear time series processes, as is the case for the AR class of models we are dealing with in this paper, as shown in the previous section, it was further proved by Kock (2016) and Medeiros and Mendes (2016), among others, that the adaptive lasso satisfies the oracle properties as well.[‡] The

[‡]Similar oracle properties have also been shown for multivariate VAR processes by Kock and Callot (2015) and Callot and Kock (2014).

adaptive lasso is therefore a theoretically sound model selection device to investigate the time-varying behavior of the optimal lag structure in realized volatility dynamics and consequently to check the validity of the HAR lag structure assumption. Note, however, that although the adaptive lasso is able asymptotically to correctly identify the non-zero coefficients, it is not really designed to check the validity of the parameter restrictions implied by the HAR model (that is, the same value of the coefficients in the three different trading windows). For this purpose, alternatives to the adaptive lasso like the clustered lasso or the group lasso can be used (see, for example, Bühlmann et al., 2013).

2.3. Hypotheses of interest and testing procedure

In our application, we are interested in testing whether the estimated adaptive lasso coefficients in an unrestricted high-order AR model are significantly different than zero. In particular, there are two hypotheses of interest related to the validity of the HAR lag structure and maximum order lag assumptions on which we will focus: The first hypothesis the HAR model implies is that all lags up to the 22nd matter for prediction and are therefore non-zero. The first corresponding null hypothesis is therefore $H_0 : \beta_i = 0, \forall i = 1, ..., 22$; a rejection of the null hypothesis is thus a confirmation of the HAR model. The second implication of the HAR model would be that lags beyond the 22nd should not matter for predicting future realized volatility, that is $H_0 : \beta_j = 0$ for $j > 22$, which would be rejected if the corresponding coefficient beyond the 22nd were found to be significantly different from zero, thus casting doubt on the correctness of the HAR dynamics. Given that we are not considering the parameter restrictions imposed by the HAR model, both tests are indirect tests for the validity of the HAR model. Constructing direct tests that consider under the alternative hypothesis the full HAR specification instead of an unrestricted AR model significantly complicates the testing procedure and is left for future research.

In order to test these hypotheses, the individual test statistics are constructed using the results of Audrino and Camponovo (2015). They recently developed a conservative inference procedure for the coefficients of the adaptive lasso. While this was previously only possible for the truly non-zero coefficients, this newly developed theory allows inference on all coefficients, thus allowing in particular to test for false positives. As we show empirically below, the presence of false positives cannot be neglected and may cause misleading interpretations of the results. Briefly, the inference procedure introduced in Audrino and Camponovo (2015) is motivated by the fact that the adaptive lasso coincides with OLS when $\lambda = 0$. The authors showed that the standard errors of the limiting distribution of the adaptive lasso estimators for each fixed value of λ become smaller when λ becomes greater than zero, which is why a conservative test for the adaptive lasso coefficients can be conducted by using the estimated OLS standard errors of the coefficients. Put formally, the individual significance of $\hat{\beta}_i$ can be found using the test statistic

$$\mathrm{T}_i = \sqrt{n}\left(\hat{\beta}_{AL,i} - \beta_i\right), \quad \forall i = 1, ..., I,\ I >> 22, \tag{6}$$

with the OLS standard errors. Further details on the implementation and asymptotic properties of the testing procedure are given in the Appendix.

2.4. Multiple hypothesis testing

It is clear that the two hypotheses introduced above can be seen as joint tests. In particular, the second hypothesis can be rewritten as $H_0 : \beta_{23} = \beta_{24} = ... = 0$ which would be rejected if one of

the coefficients beyond 22 were found to be significantly different from zero, thus implying that the maximal order lag assumed under an HAR model specification were wrong. Practically, of course, only a finite number of lags can be tested but at the same time the order of the AR model should be chosen sufficiently large. For the sake of the exposition, let us assume that the first 100 lags will be considered (that is, $I = 100$ in (6)). In that case, the first hypothesis would be equivalent to a joint test of 78 individual hypotheses. This large number of hypotheses for the joint test justifies the use of a procedure that corrects for the fact that we are dealing with a multiple testing problem. Among those introduced in the literature[§], we consider the procedure that controls the *family-wise error rate*, as suggested by Holm (1979) and Romano and Wolf (2005).

The intuition behind this procedure is that for a large number of hypotheses one would expect to reject some test statistics in any case. The Holm procedure corrects for this fact by imposing a threshold on the probability α that one or more of the hypotheses are rejected. To state the exact procedure, we assume that the p-values of the individual s statistics are sorted in ascending order, so that $\hat{p}_1 \leq \hat{p}_2 \leq ... \leq \hat{p}_s$. The step-wise procedure is then given by (Romano and Wolf, 2005):

1. If $\hat{p}_1 \geq \alpha/s$, accept $H_1, ...H_s$ and stop. If $\hat{p}_1 < \alpha/s$, reject H_1 and test the remaining $s - 1$ hypotheses at level $\alpha/(s - 1)$.
2. If $\hat{p}_1 < \alpha/s$ but $\hat{p}_2 \geq \alpha/(s - 1)$, accept $H_2, ...H_s$ and stop. If $\hat{p}_1 < \alpha/s$ and $\hat{p}_2 < \alpha/(s - 1)$, reject H_2 in addition to H_1 and test the remaining $s - 2$ hypotheses at level $\alpha/(s - 2)$.
3. And so on.

It needs to be stressed that this procedure is very conservative because for large numbers of hypotheses, the critical values will become very large. This would make the test tend to view the coefficients beyond the 22nd as not significantly different from zero jointly. Given that the individual test statistics are already conservative, this test could therefore be considered as too conservative in the case at hand.

3. Empirical Analysis

3.1. Data set

We consider tick prices obtained from TickData consisting of intraday quotes of Alcoa, American Express, Baxter, Blackrock, Citigroup, Coca Cola, Dow, Exxon Mobil, Gilead, Goldman Sachs, Hasbro, Harley Davidson, Intel Corporation, Met, Microsoft, Nike, Pfizer, Verizon and Yahoo from January 2, 2001 to November 15, 2010, which corresponds to 2472 trading days after filtering.

The intraday quotes were used to compute daily realized variance measures based on the two-time scales estimator by Zhang et al. (2005), given its robustness in the presence of microstructure noise.[¶] The two chosen scales were 2 and 20 ticks; however, the results are not sensitive to the grid size. In order to further mitigate the effect of microstructure noise, we also disregard the trades of the first and last 30 minutes of each trading day.

[§]See, as alternatives, Benjamini and Hochberg (1995) or Genovese et al. (2006).

[¶]We found that our results do not change when using another estimator that is robust to microstructure noise, e.g. a realized kernel. In the presence of jumps, using the two-time scales estimator we end up focusing on the dynamics of *total* quadratic variation (that is, including the contribution of squared jumps). Instead, if the interest is on the continuous part of quadratic variation, estimators that are robust to jumps, like the MedRV estimator, can be considered.

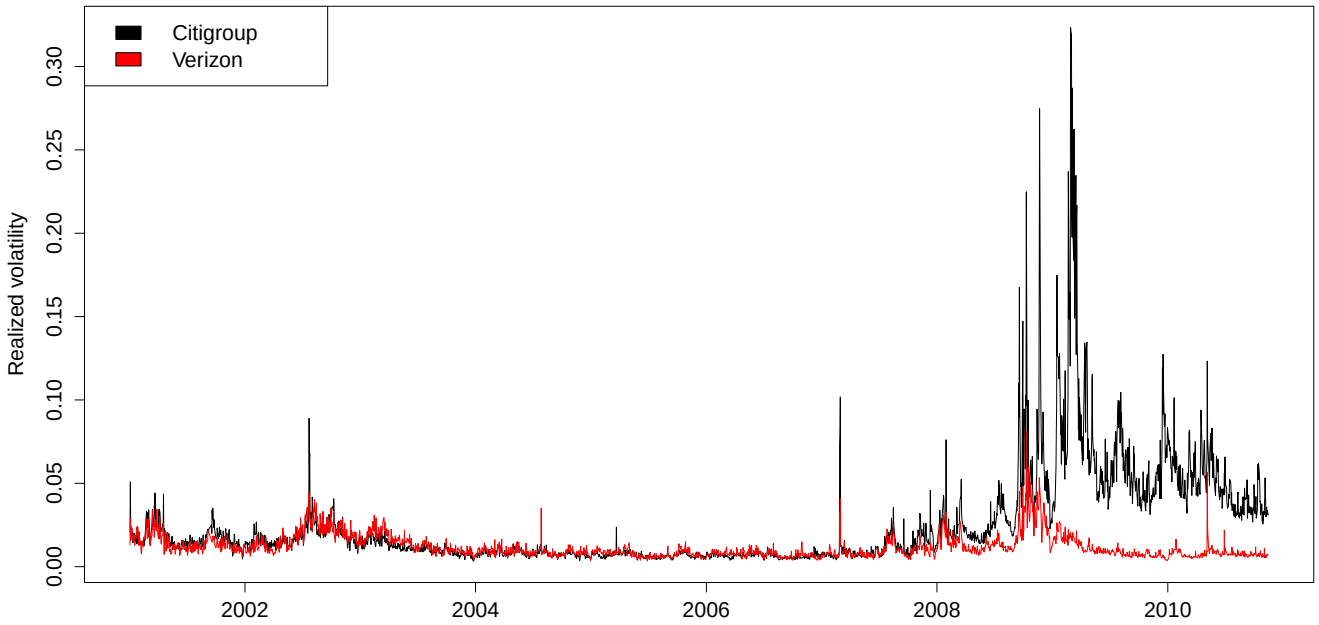

Figure 1. Time series of realized volatility for Citigroup (black) and Verizon (red) for the time period from January 2, 2001, to November 15, 2010.

Figure 1 shows some illustrative examples of the computed realized volatility series for Citigroup and Verizon, belonging to the financial and technological sector, respectively.[‖] It can be seen that, in general, realized volatility levels are higher for Citigroup than for Verizon. This is in line with expectations given the respective sectors. Summary statistics for the series under investigation are given in Table 1.

Furthermore, in the considered time period we can identify two main events affecting realized volatility for all assets. The first is the period until roughly the fourth quarter of 2002, which corresponds to the aftermath of the burst of the tech bubble and the terrorist attacks on 9/11. In this period, realized volatility levels were higher than usual for all assets and more spikes can be observed. The subsequent period from 2003 to the end of 2007 can be viewed as relatively calm, with realized volatility levels remaining relatively constant with only a few spikes. The third period starts in 2008, with realized volatility reaching extremely high levels during the start of the financial crisis. It should be noted that in this period realized volatility returns to pre-crisis levels in 2009 for most non-financial assets, while Citigroup and other financial assets continue to have higher than usual levels of volatility. The mentioned properties can be observed across all assets, with only Citigroup being an exception with a persistently high realized volatility during the financial crisis.

[‖]The sector classifications for all assets were adopted from Yahoo Finance.

Table 1. Descriptive statistics of the realized volatility series for all assets under investigation. The top panel contains statistics for the entire time period from January 2, 2001 to November 15, 2010, the second panel for subperiod 1 up to the end of the US stock market downturn in 2002 (2002/10/09), the third for subperiod 2 between the end of the stock market downturn and the collapse of Lehman Brothers (2008/09/15), and the fourth for the final subperiod after the collapse of Lehman Brothers.

		AA	AXP	BAX	BLK	C	DOW	GILD	GS	HAS	HOG
Total	Mean	168.84	149.89	108.87	132.66	244.08	143.78	170.91	153.32	133.06	146.72
	SD	98.14	111.32	56.70	93.98	293.89	81.38	85.58	119.87	75.83	86.31
	Min	49.38	22.77	29.11	0.41	33.06	33.49	43.11	31.08	18.33	35.92
	Max	1405.28	1219.51	1016.44	771.15	3233.65	926.15	1394.36	2118.57	776.32	748.34
Period 1	Mean	181.13	194.03	137.94	116.12	187.50	181.89	279.92	175.96	201.22	169.23
	SD	67.66	75.82	75.77	90.55	80.21	72.95	95.02	73.92	91.70	69.87
	Min	65.63	52.51	29.11	0.41	45.59	51.11	100.49	46.15	62.65	42.84
	Max	442.67	565.51	1016.44	562.54	889.96	696.40	748.41	961.02	776.32	615.55
Period 2	Mean	134.37	109.57	99.27	113.67	113.26	110.62	151.60	124.59	107.43	110.01
	SD	56.33	71.15	41.51	61.96	80.67	46.59	41.91	63.66	47.31	45.18
	Min	49.38	22.77	34.35	6.67	33.06	33.49	43.11	31.08	18.33	35.92
	Max	952.57	635.93	475.88	438.40	1017.87	498.73	436.94	678.88	376.38	372.56
Period 3	Mean	252.99	224.85	112.08	197.38	644.88	204.01	137.54	213.70	149.10	228.93
	SD	144.21	161.40	66.12	132.58	396.51	109.67	98.30	207.97	86.78	115.74
	Min	90.21	49.21	33.22	51.73	227.66	72.96	44.43	49.23	41.94	66.33
	Max	1405.28	1219.51	597.71	771.15	3233.65	926.15	1394.36	2118.57	620.93	748.34

		INTC	KO	MET	MSFT	NKE	PFE	VZ	XOM	YHOO
Total	Mean	157.83	94.48	154.49	125.83	122.88	121.24	118.20	114.61	188.35
	SD	73.93	50.32	124.83	68.12	65.37	60.13	69.22	71.00	104.29
	Min	55.12	22.80	22.64	35.77	37.18	44.87	34.16	29.32	52.27
	Max	1017.91	627.35	1086.87	1325.89	656.84	991.75	811.64	1420.57	1501.75
Period 1	Mean	228.77	134.83	164.05	171.64	159.84	148.56	162.33	138.94	333.43
	SD	85.43	51.49	66.52	64.65	62.66	55.59	70.84	56.54	93.63
	Min	64.84	49.47	55.83	48.98	42.35	72.28	60.00	41.65	155.58
	Max	982.62	404.98	484.59	748.56	387.83	521.40	445.49	387.19	713.50
Period 2	Mean	129.57	80.88	111.07	102.66	101.43	98.54	103.61	101.88	149.65
	SD	38.56	31.11	51.87	45.20	43.11	40.25	47.97	43.33	56.17
	Min	55.12	22.80	22.64	45.18	37.18	44.87	34.16	29.32	52.27
	Max	468.39	406.00	491.75	1325.89	318.68	789.05	407.87	599.81	672.48
Period 3	Mean	178.88	99.70	265.18	152.80	152.15	161.50	123.13	130.10	179.32
	SD	91.05	69.96	205.15	90.81	89.18	76.56	96.31	118.25	113.30
	Min	55.23	28.48	49.34	35.77	46.39	59.69	36.09	30.74	63.33
	Max	1017.91	627.35	1086.87	952.95	656.84	991.75	811.64	1420.57	1501.75

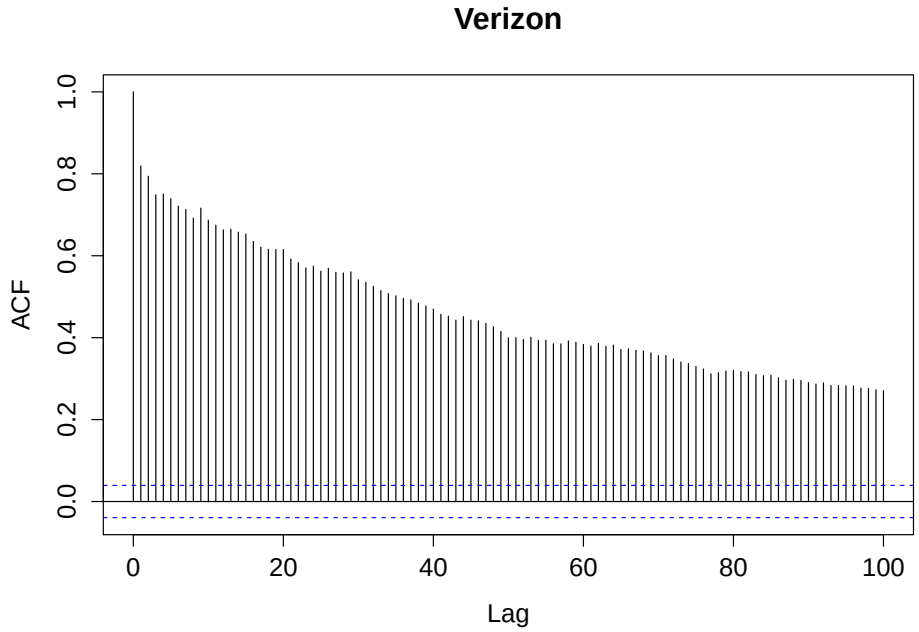

Figure 2. Autocorrelation function of realized volatility for Verizon up to lag 100.

Figure 2 shows an example of the autocorrelation function for the time series of realized volatility of Verizon. This demonstrates that the autocorrelation function is very slowly decaying and confirms the well documented long-memory shown by realized volatility series. Moreover, it also provides clear evidence for the high correlation present among near lags of realized volatility. From an estimation point of view, this will lead to highly collinear regressors in our AR setting, which makes the isolation of the individual effects more difficult. This point must be carefully taken into account when discussing the results of the adaptive lasso analysis, in particular those related to the validity of the HAR lag structure when using the first testing hypothesis introduced in Section 2.3.

3.2. Rolling window analysis

The empirical analysis is conducted using a rolling window procedure to overcome problems related to the possible non-stationarity of the realized volatility series: see the discussion after Figure 1 and the extensive literature about structural breaks in realized volatility series such as Hillebrand and Medeiros (2016) and Bertram et al. (2013) and the references therein. The first step is the estimation of the adaptive weights λ_i. In this case, the OLS estimator is chosen in accordance with Audrino and Camponovo (2015), although theoretically any $\sqrt{n}$-consistent estimator would be a valid choice as well.[**]

Estimating the regression of the time series on its first $I = 100$ lags by OLS, the resulting coefficients are inverted in absolute value to obtain the weights:

$$\lambda_i = 1/|\hat{\beta}_{OLS,i}|, \quad i = 1, \ldots, 100. \tag{7}$$

[**]Empirically, it turns out that the choice of this first-step estimator is of very little importance and produces almost identical final results, as for example from using the standard lasso or other valid estimators.

These weights are then used to fit an adaptive lasso to the model, which of course still depends on the tuning parameter λ.

We choose λ based on the one-standard-error rule via cross-validation. All estimations are conducted in R via the glmnet package: see Friedman et al.(2010)[††]

3.2.1. Adaptive lasso estimation

In a first step, we will analyze which lags are selected by employing the adaptive lasso. The above-mentioned procedure was conducted on a rolling window basis, with a window length of 1000 trading days (that is, about 4 years), rolled over day by day.[‡‡]The top panels in Figures 3 through 5 show the lags that are selected (blue rectangles) by the adaptive lasso for our illustrative examples, namely Citigroup, Verizon and Exxon Mobil. Results for the rest of the assets can be obtained from the authors upon request.

Each column corresponds to the selected lags for one rolling window, with the start date and end date given at the top and bottom of the plot, respectively. Additionally, we mark four events in the timeline and the dates at which they left (entered) the rolling windows: 9/11, the end of the stock market downturn (we choose the day when the S&P 500 reached its lowest level), and the collapse of Bear Stearns (BS) and Lehman Brothers (LB). The diagonal dotted lines have slope one and help to illustrate how individual events propagate through the lag structure over time. The red line indicates the maximal order lag implied by the HAR model assumption.

Inspecting the plots of the results, we can make three key observations: First, we see that generally not all of the first 22 lags are selected by the adaptive lasso. This confirms the findings by Audrino and Knaus (2016). In fact, usually no more than 5 or 6 lags among the first 22 are selected by the adaptive lasso. Moreover, generally these active lags are among the first ten. This contradicts the implications of the HAR model. However, given the high multicollinearity of the lags of realized variance, it needs to be emphasized that this result should be viewed with caution: Since the lags are highly correlated and given that the adaptive lasso is not enforcing the parameter restrictions implied by the HAR model, it is to be expected that the estimation procedure will select only a few of these lags and that these will be highly significant. In other words, due to the penalization of any non-zero coefficient, it is not surprising that the adaptive lasso selects only a few of these lags that seem to have a highly similar informational content. Thus, the fact that the adaptive lasso does not select all the necessary lags should not be viewed as strong evidence against the validity of the HAR model.

Second, we see that the adaptive lasso selects many lags beyond the 22nd for many assets. In fact, it sometimes selects lags that are considerably far beyond the 22nd, for example for Citigroup and Verizon. These lags are not in line with the implications of the HAR model and furthermore appear to be relatively stable over long periods of time. This, as well, confirms the results by Audrino and Knaus (2016).

The third observation that can be made by inspecting the rolling window plots is that two structural breaks are clearly visible across almost all assets which divide the time period under investigation into three subperiods. It appears that the end of the stock market downturn and the collapse of Lehman

[††]Alternative ways of finding λ, for example based on the minimization of the BIC, have proved to be very similar after testing. Before testing, however, the adaptive lasso generally selects many more variables when minimizing the BIC than when applying the one-standard-error rule. We interpret this result as a signal that optimizing the shrinkage parameter using the BIC generally leads to many more false positives than when using cross-validation.

[‡‡]We find that choosing a different window length does not qualitatively alter the main results and conclusions.

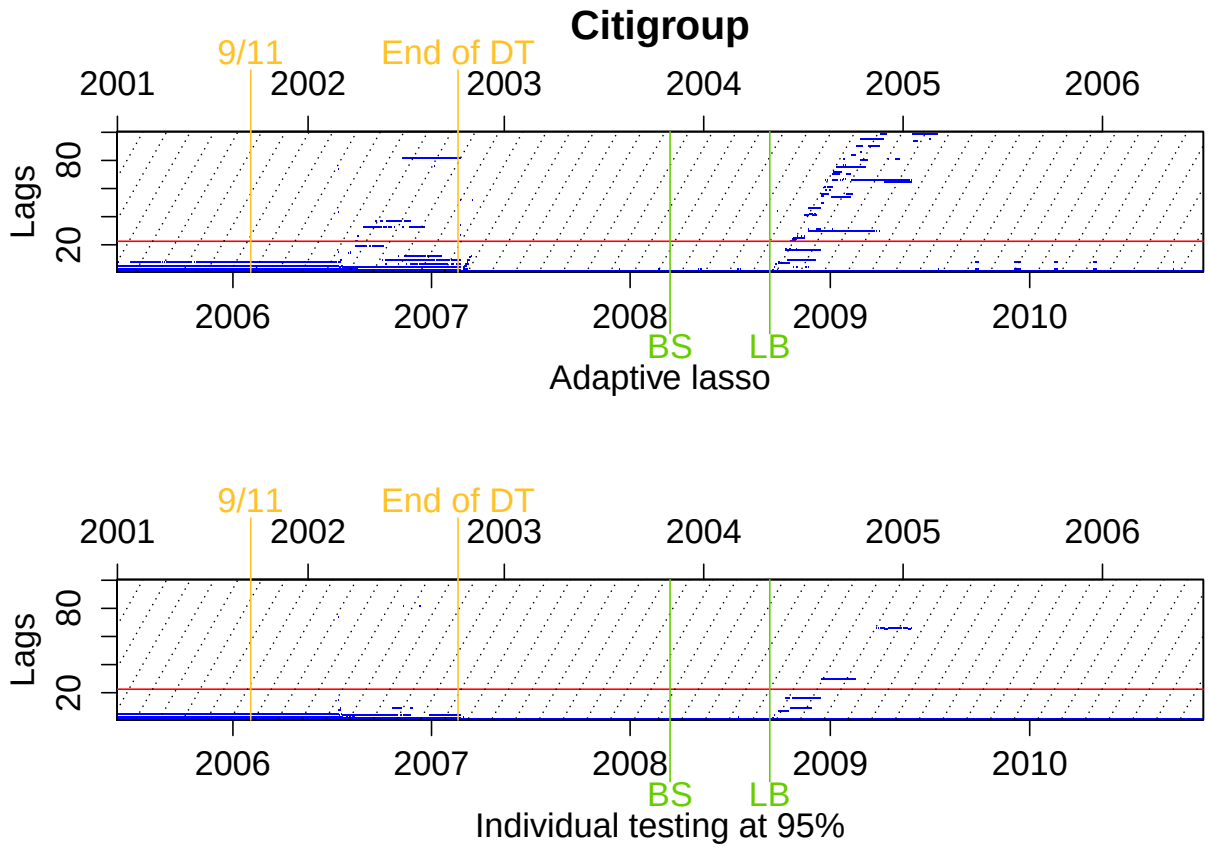

Figure 3. Lags selected by the adaptive lasso (top) and lags significant at the 95% significance level (bottom) for Citigroup: The blue rectangles correspond to the lags selected by the adaptive lasso. The top (bottom) axis contains the start (end) dates of the respective rolling windows. The red line is drawn after the 22nd lag. Additionally, four events are highlighted: 9/11, the end of the US stock market downturn in 2002, and the collapse of Bear Stearns (BS) and Lehman Brothers (LB).

Brothers mark two break points at which the lag structure selected by the adaptive lasso significantly changes. The effect of the collapse of Lehman Brothers is best visible by inspecting the lags selected for Citigroup in Figure 3: In the third period, right after the collapse and along the diagonal lines, the shocks to volatility propagate through the lag structure and remain for a short period. This pattern is, not surprisingly, also easily visible for the other financial titles.

The breakpoint dividing the first and the second subperiod appears to be the end of the stock market downturn during the final months of 2002. This is in line with the results of Hillebrand and Medeiros (2016) and Audrino and Knaus (2016) who also confirm the presence of a structural break in realized volatility around the end of 2002. In general, across many assets, it can be observed that larger lags tend to be selected as active variables by the adaptive lasso more often in the first and the third period than in the second. Throughout the second period there is very little variation in the lags selected, which is not surprising given that this period is characterized by calm market conditions. Furthermore,

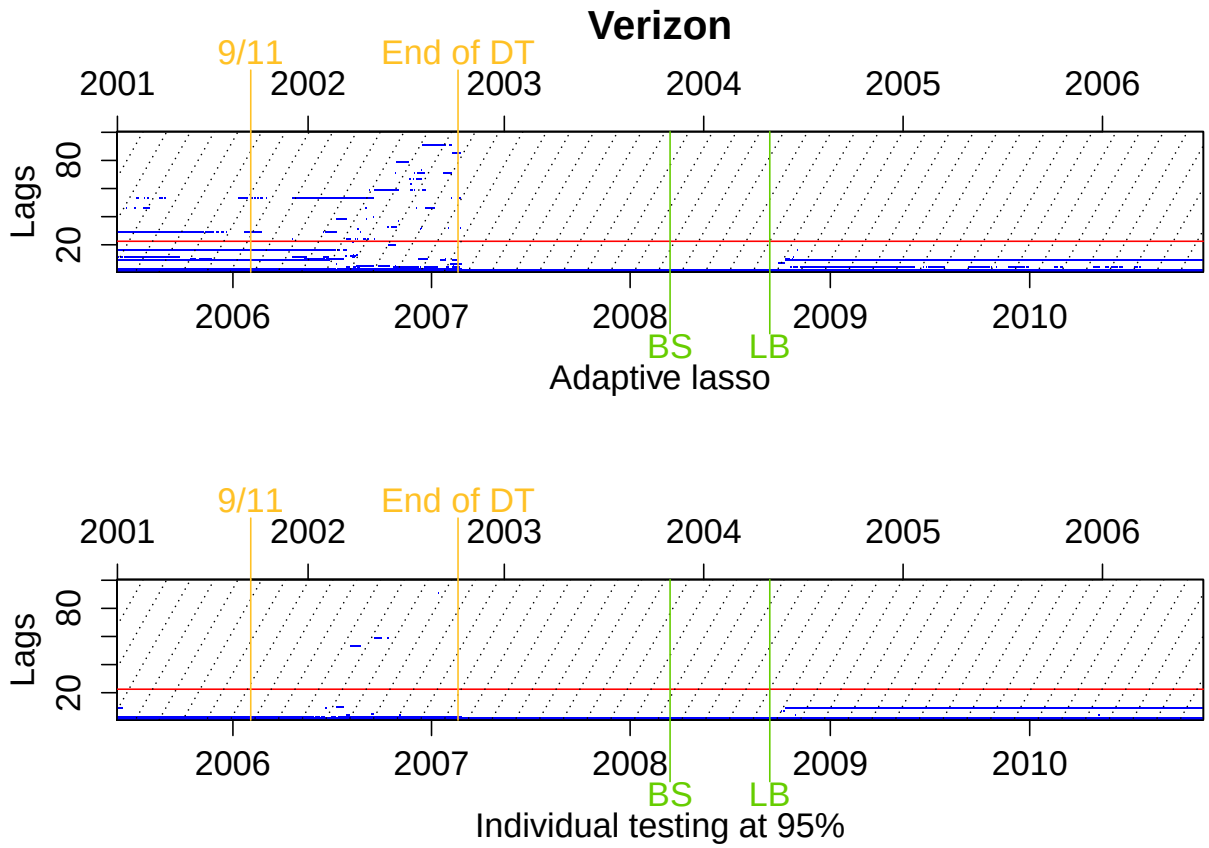

Figure 4. Lags selected by the adaptive lasso (top) and lags significant at the 95% significance level (bottom) for Verizon: The blue rectangles correspond to the lags selected by the adaptive lasso. The top (bottom) axis contains the start (end) dates of the respective rolling windows. The red line is drawn after the 22nd lag. Additionally, four events are highlighted: 9/11, the end of the US stock market downturn in 2002, and the collapse of Bear Stearns (BS) and Lehman Brothers (LB).

across assets the adaptive lasso selects only very few and very short lags in this period: for many assets, in fact, only up to the third lag. In other terms, it seems that during such a quiet period, characterized by low levels of realized volatility, very simple autoregressive processes perform best at predicting volatility. In fact, recalling the plot of realized volatility series during the 2003-2007 subperiod, we see that the persistence of the series is significantly lower than that observed in the preceding and following periods. The more quickly decreasing autocorrelation function in that period is therefore in line with the smaller order of the autoregressive process.

A plausible explanation for the fact that in this calm period only the shortest lags are selected is that there are far fewer spikes of volatility observable than in more hectic times, such as the first and the third period: The second period is characterized by a relatively constant level of volatility, which means that only very few events occur which affect volatility. Therefore, the time series of volatility is very smooth and given the lack of events with big impact, the preceding day's level of volatility

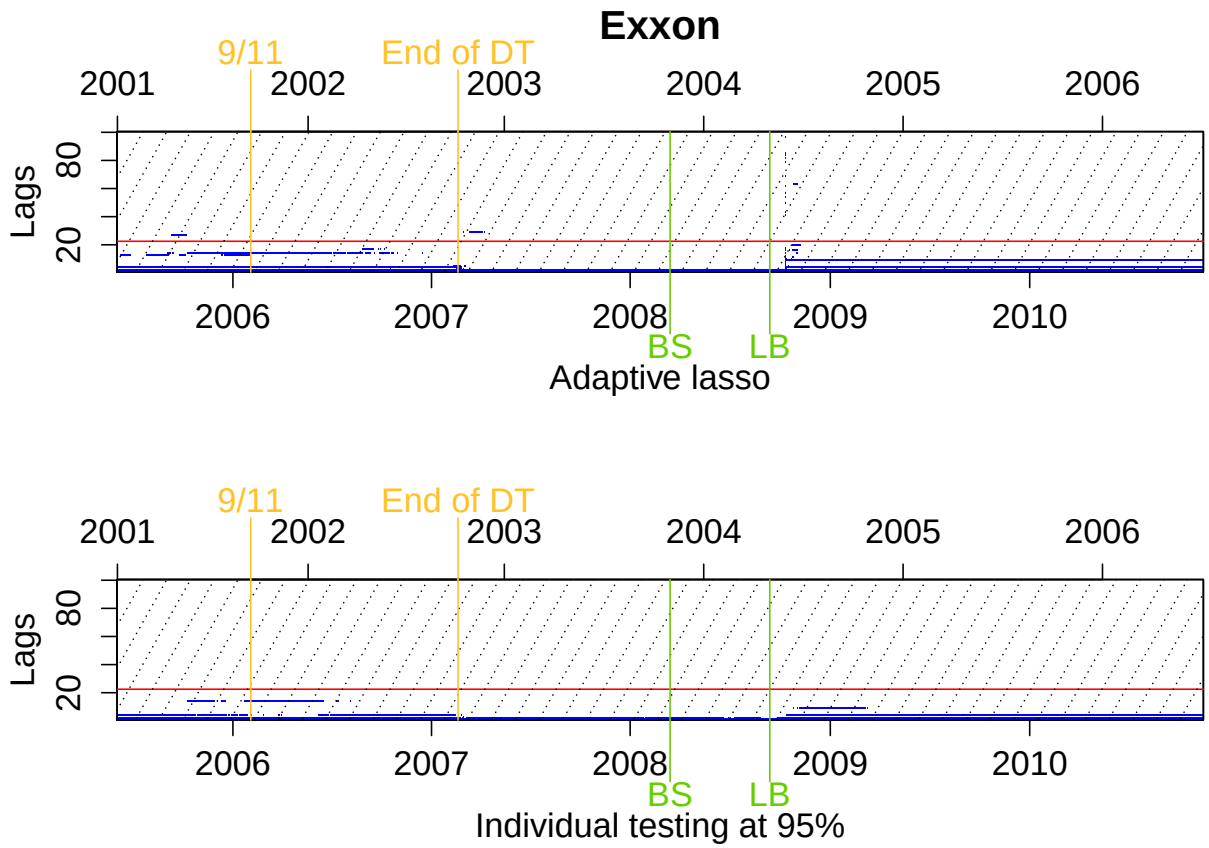

Figure 5. Lags selected by the adaptive lasso (top) and lags significant at the 95% significance level (bottom) for Exxon: The blue rectangles correspond to the lags selected by the adaptive lasso. The top (bottom) axis contains the start (end) dates of the respective rolling windows. The red line is drawn after the 22nd lag. Additionally, four events are highlighted: 9/11, the end of the US stock market downturn in 2002, and the collapse of Bear Stearns (BS) and Lehman Brothers (LB).

is a high quality forecast. On the other hand, for the more hectic periods, there are many events that occur at a high frequency which lead to volatility spikes. Given the persistence of volatility in these periods, longer lags therefore have much greater forecasting power. With respect to the high number of volatility spikes, it is thus intuitively compelling that larger lags are also selected by the adaptive lasso, as compared to the calmer period.

Fourth, while this general pattern seems to hold across most of the assets in our sample, we also detect two other factors that have an impact on the lags selected by the adaptive lasso in the different subperiods: the market sector and the beta of the individual assets. Figure 6 shows how the adaptive lasso estimates for the financial stocks compare to those for the stocks of the basic materials sector. We can observe that generally many more large lags are selected for the financial assets than for the basic materials assets, across all periods. An intuitive explanation for this behavior is that compared to financial companies, basic materials companies are less subject to particular financial and/or

economic events that have a big impact on volatility. Therefore, volatility progresses more smoothly and the shorter lags are much more valuable for forecasting than the large ones. On the other hand, the larger lags selected for the financial stocks, in particular those that propagate over the rolling windows, usually identify particular events yielding large spikes in realized volatility. For example, we see that the collapse of Lehman Brothers visibly propagates through the lag structure of the financial assets but does not have any effect on the basic materials assets.

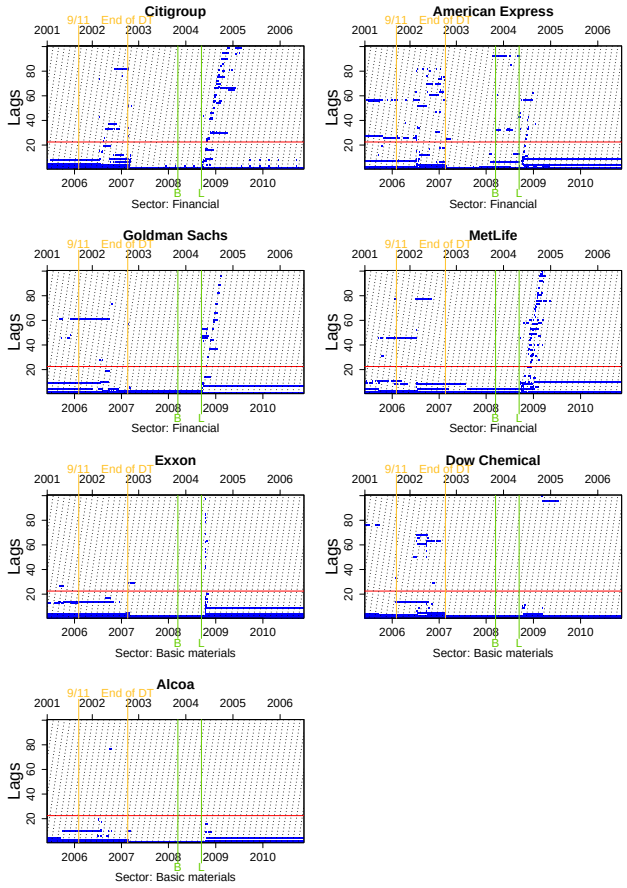

Figure 6. Lags selected by the adaptive lasso for financial (top) and basic materials (bottom) assets: The blue rectangles correspond to the lags selected by the adaptive lasso. The top (bottom) axis contains the start (end) dates of the respective rolling windows. The red line is drawn after the 22nd lag. Additionally, four events are highlighted: 9/11, the end of the US stock market downturn in 2002, and the collapse of Bear Stearns (BS) and Lehman Brothers (LB).

A second relevant factor impacting the lag structure selected by the adaptive lasso is the asset's beta. Figure 7 shows the grouped plots of the adaptive lasso estimates of the four assets with the highest betas (top) and the four assets with the lowest betas (bottom). In order to isolate the effect of beta from the effect of the sector, the plots show the assets with the highest and lowest betas from different sectors. We can see from these plots that the third period of the financial crisis is actually not only related to the sector but also to the beta. For the high beta assets we observe that lags beyond the 22nd are selected

1981 times in total in that period (in contrast to only 18 times for the low beta assets). This can again be explained by the fact that low beta assets are less influenced by market turbulence and therefore show fewer significant spikes of volatility in that period. While we can see that the sector and the beta have an influence on the lag structure that is selected by the adaptive lasso, we find that other factors such as liquidity and value do not seem to show visible effects.

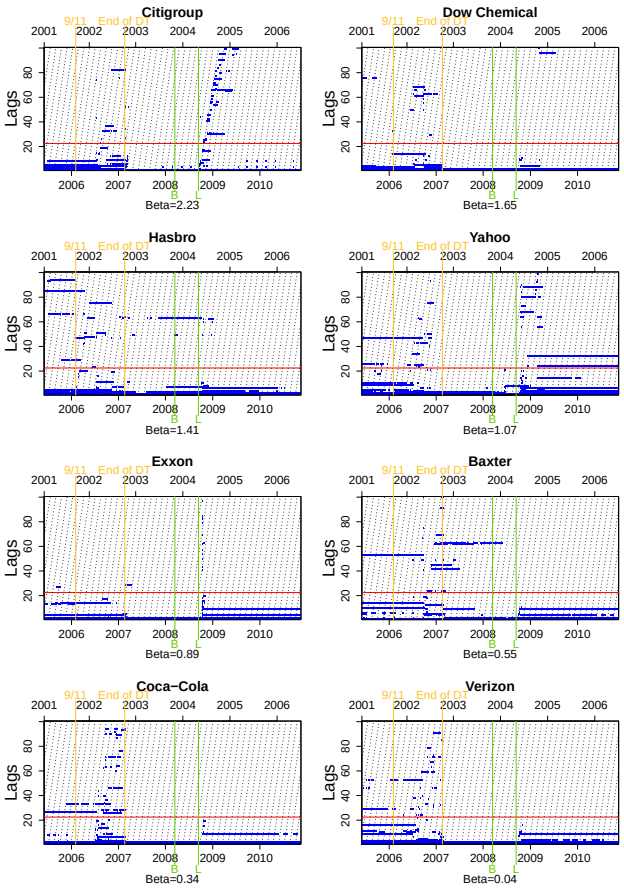

Figure 7. Lags selected by the adaptive lasso for high beta (top) and low beta (bottom) assets from different sectors: The blue rectangles correspond to the lags selected by the adaptive lasso. The top (bottom) axis contains the start (end) dates of the respective rolling windows. The red line is drawn after the 22nd lag. Additionally, four events are highlighted: 9/11, the end of the US stock market downturn in 2002, and the collapse of Bear Stearns (BS) and Lehman Brothers (LB).

3.2.2. Testing for false positives

Having obtained the adaptive lasso estimates, we now test to see whether these estimates are indeed significantly different from zero or can be classified as false positives. We therefore test the two hypotheses of interest using the procedure described in section 2. This is done by dividing each estimate $\hat{\beta}_{AL,i}$ by its respective OLS standard error that was previously computed. The absolute value of the resulting quantity is then compared to the normal quantiles at the desired significance level; in

our case we test at confidence level 95%. The plots reporting the significantly selected lags after testing are drawn in the bottom panels of Figures 3 through 5 for the three illustrative examples Citigroup, Verizon and Exxon Mobil, respectively.

Analyzing these plots, we can revisit the discussion of the previous subsection after controlling for false positives (although in a conservative way). To begin with, the first hypothesis is generally rejected because not all of the first 22 lags are selected by the adaptive lasso. We now see that most of these lags are also found to be significant, with some of the very short lags being the most significant. We can therefore conclude that much of the information contained in the lag structure is in fact due to the shorter lags. Nevertheless, similarly to what we discussed above, given that short lags are highly correlated, this result cannot be seen as strong evidence against the HAR model.

Second, while the adaptive lasso estimates seem to reject the second hypothesis, we now see that after testing we indeed fail to reject it at a significance level of 95% for most assets at most times. After testing, we observe that almost all lags that are selected beyond the 22nd are insignificant at the 95% level and can be generally interpreted as false positives. Except for a very few very short periods of time, the lags disappear completely, which therefore supports the validity of the HAR model. In fact we see that only four of the 19 assets retain a large lag that remains significant for a long period. On the other hand, however, given that we are using a conservative procedure favoring the null hypothesis of no active lags beyond 22, those exceptions can be interpreted as strong evidence against the lag structure and maximal order lag implied by the HAR model for the corresponding assets and subperiods.

Overall, the results regarding the significant lags provide a very good explanation as to why the HAR model performs so well in practice in the different forecasting applications: usually all the lags that are significantly selected by the adaptive lasso for forecasting tomorrow's realized variance are among the first 22 lags. The subset of lags that the HAR model focuses on therefore appears to contain all the relevant information. Due to the fact that these lags are highly correlated, it is also not as important which of these lags are used. Thus, it does not make a crucial difference that the HAR model, as opposed to the adaptive lasso, focuses on all of the first 22 lags. Since the HAR model imposes restrictions on the estimated parameters, it is nonetheless very parsimonious. This finding also explains the results found by Craioveanu and Hillebrand (2012) who show that the choice of the aggregation frequencies does not make a significant difference for the performance of the HAR model.

Finally, we see that while most of the large lags are found to be insignificant, we still observe two breaks in structure among the significant ones. This is best seen by acknowledging the fact that the short lags that are significantly selected also usually change when the breaks occur.

3.2.3. Multiple testing

As mentioned in the theoretical section, the second hypothesis amounts to a joint test of 78 individual hypotheses. This would justify the procedure of controlling the family-wise error rate instead of the individual ones. We report the results for all assets for a significance level of 95% in Figure 8.

For the lags beyond the 22nd, the plots report the lags that are significant after multiple testing, while for the lags up to the 22nd, they report the ones that are significant after individual testing. Inspecting the figures, we see that after multiple testing all the lags beyond the 22nd are insignificant for most assets most of the time. One exception would be Microsoft, for example, for which a large

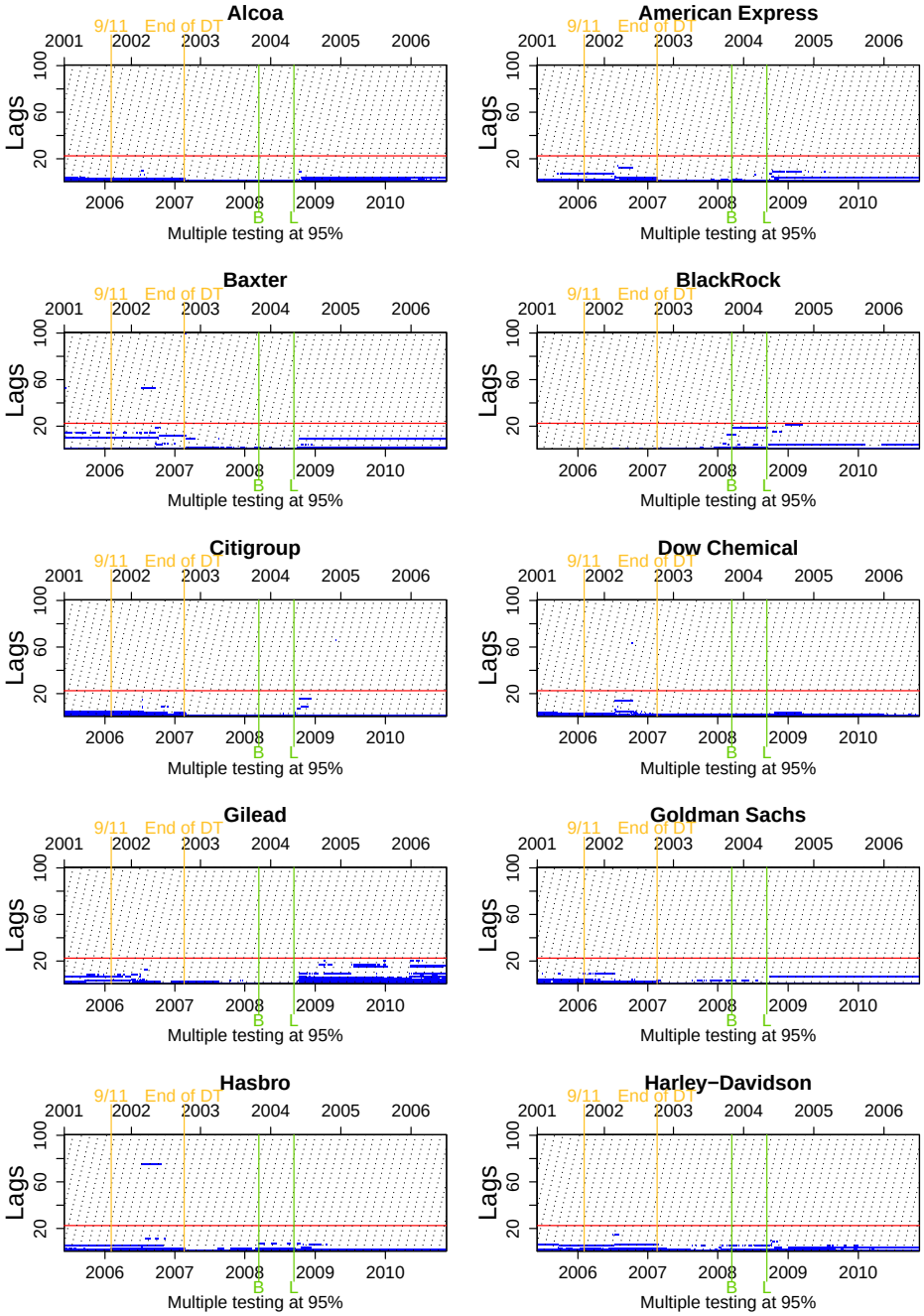

Figure 8. Lags that are significantly selected after testing the lags beyond the 22nd multiply at the 95% significance level: The blue rectangles correspond to the lags selected by the adaptive lasso. The top (bottom) axis contains the start (end) dates of the respective rolling windows. The red line is drawn after the 22nd lag. Additionally, four events are highlighted: 9/11, the end of the US stock market downturn in 2002, and the collapse of Bear Stearns (BS) and Lehman Brothers (LB).

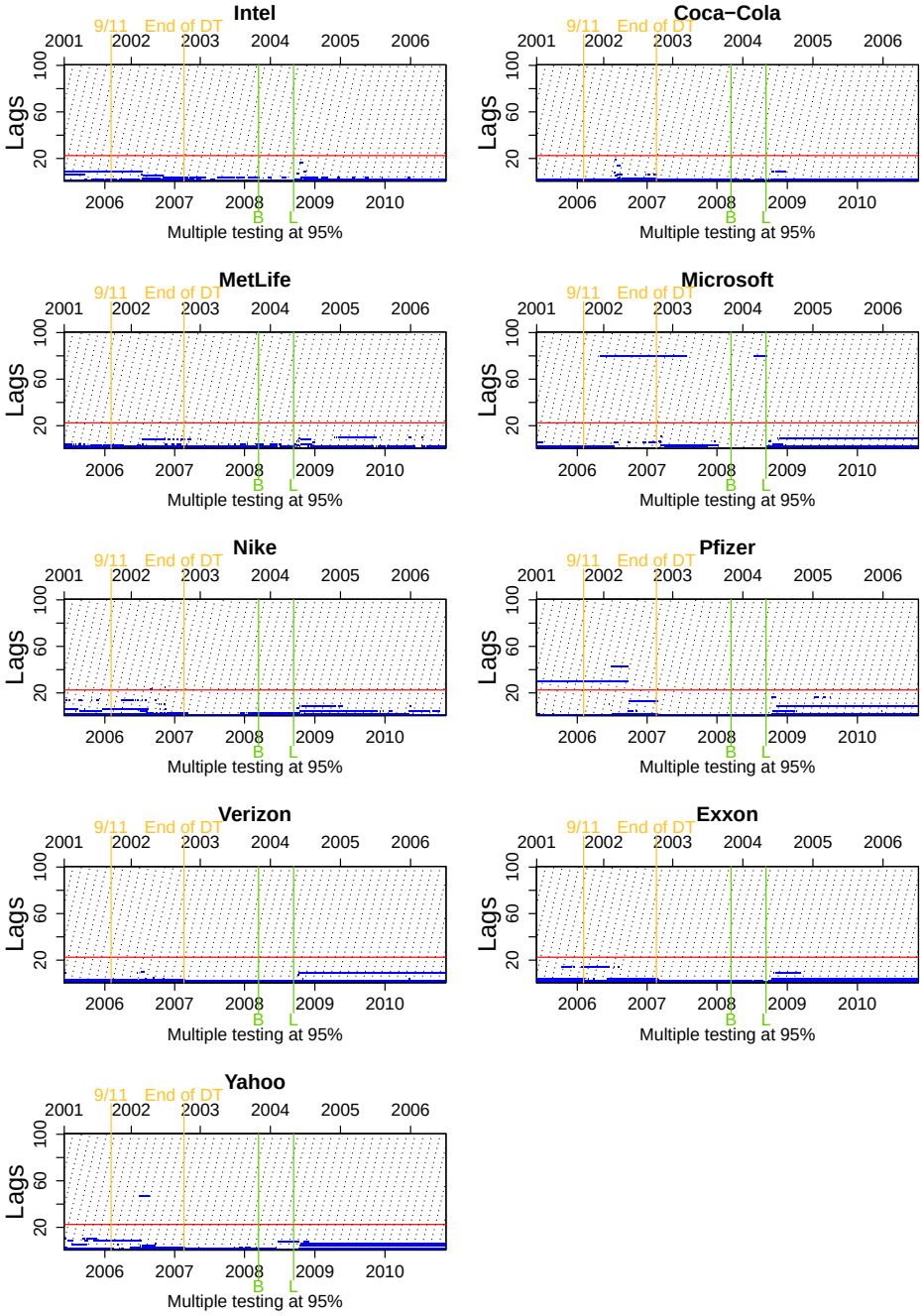

Figure 8. (continued)

lag is significantly selected throughout the first and second period. It is not surprising that generally most of the large lags turn out to be insignificant after testing multiply, given that the test procedure, as previously mentioned, is very conservative. Together with the already conservative way of estimating the standard errors, this amounts to a test which is too conservative in the sense that the critical values for the lags become extremely high.

4. Conclusion

The goal of this paper was to estimate the optimal lag structure driving the behavior of the realized volatility dynamics in an autoregressive setting. Moreover, given that the HAR model assumptions are equivalent to stating that the dynamics of realized volatility follows a constrained AR (22) process, we also checked in an indirect way whether the structural assumptions that the HAR model implies can be recovered using solely statistical techniques.

For the empirical investigation, we employ the adaptive lasso as a selection device for the past lags that are most relevant for prediction. We make the observation that the lags selected for the 19 stocks under investigation generally do not corroborate the validity of the lag structure and maximal order lag implied by the HAR model: For most assets, lags far beyond the 22nd are selected and not all of the first 22 are selected either.

We further observe in the rolling window analysis that the lags selected are generally relatively stable over time. However, there appear to be two structural breaks, which significantly affect the lag structure across all assets, with the first one occurring at the end of the stock market downturn at the end of 2002 and the second one coinciding with the collapse of Lehman Brothers in 2008. We further see that for almost all assets, during the middle period, which is very calm in terms of realized volatility, only very short lags are selected by the adaptive lasso. This is in contrast to the more volatile periods in the beginning and the end in which for some assets very large lags are selected additionally. A likely explanation for this observation is that the larger lags are related to the frequency of spikes of volatility: In calm periods, a simple autoregressive process of small order is a very good predictor of the future behavior because volatility progresses very smoothly. For the more volatile periods in the beginning and in the end, large spikes of volatility cause the adaptive lasso to attribute predictive power even to very large lags.

Another observation that can be made is that the sector and the beta seem to have an effect on how the lag structure is affected by the structural breaks. Comparing the results of financial assets with basic materials assets, we see that for basic materials assets, significantly fewer large lags are generally selected. This can again be attributed to the fact that fewer spikes of volatility occur for these assets. Furthermore, for the financial titles, we see that the collapse of Lehman Brothers clearly propagates through the lag structure. In fact, the effect of the financial crisis is also closely linked to the size of the beta: We see that for low beta assets, in contrast to high beta assets, even across sectors virtually no large lags are selected in the final period.

Finally, we find that for almost all assets over all time windows, the large lags that are inconsistent with the HAR model are statistically insignificant. This explains the good empirical performance of the HAR model because it shows that the model focuses on the right subset of the lags that are significant for forecasting. The fact that not all of the first 22 lags are significant should not be overrated given that the HAR model puts constraints on the autoregressive process and the fact that the first lags are

highly correlated. The HAR model therefore focuses in a parsimonious way on the subset of lags that contain the most important information.

Conflict of Interest

All authors declare no conflicts of interest in this paper.

References

Andersen T, Bollerslev T, Diebold F (2003) Some like it smooth, and some like it rough: untangling continuous and jump components in measuring, modeling, and forecasting asset return volatility. Pier working paper 03-025, Northwestern University - Kellogg School of Management.

Andersen TG, Bollerslev T, Diebold FX, et al. (2001) The distribution of realized stock return volatility. *J Financ Econo* 61: 43-76.

Andersen TG, Bollerslev T, Diebold FX, et al. (2001) The distribution of realized exchange rate volatility. *J Am Stat Associ* 96: 42-55.

Andersen TG, Bollerslev T, Diebold FX, et al. (2003) Modeling and forecasting realized volatility. *Econom* 71: 579-625.

Ang A, Bekaert G (2007) Stock return predictability: Is it there? *Rev Financ Stud* 20: 651-707.

Audrino F, Camponovo L (2015) Oracle properties, bias correction, and inference of the adaptive lasso for time series extremum estimators. SSRN working paper series, University of St. Gallen.

Audrino F, Knaus SD (2016) Lassoing the har model: A model selection perspective on realized volatility dynamics. *Econom Rev* 35: 1485-1521.

Barndorff-Nielsen OE, Shephard N (2002) Estimating quadratic variation using realized variance. *J Appl Econom* 17: 457-477.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc, Ser B* 57: 289-300.

Bertram P, Kruse R, Sibbertsen P (2013) Fractional integration versus level shifts: the case of realized asset correlations. *Stat Papers* 54: 977-991.

Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. *J Econom* 31: 307-327.

Bühlmann P, Rütimann P, Geer SVD, et al. (2013) Correlated variables in regression: clustering and sparse estimation. *J Stat Plan Inference* 143: 1835-1858.

Callot LAF, Kock AB (2014) Oracle efficient estimation and forecasting with the adaptive lasso and the adaptive group lasso in vector autoregressions. In: N. Haldrup, M. Meitz and P. Saikkonen (eds), *Essays Nonlinear Time Ser Econom*, Oxford University Press.

Campbell JY, Yogo M (2006) Efficient tests of stock return predictability. *J Financ Econ* 81: 27-60.

Corsi F (2009) A simple approximate long-memory model of realized volatility. *J Financ Econom* 7: 174-196.

Corsi F, Audrino F, Renò R (2012) HAR Modelling for Realized Volatility Forecasting. In *Volatility Models Appl*. Wiley.

Craioveanu M, Hillebrand E (2012) Why it is ok to use the HAR-RV (1, 5, 21) model. Working paper, University of Central Missouri.

Friedman JH, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 33: 1-22.

Genovese CR, Roeder K, Wasserman L (2006) False discovery control with p-value weighting. *Biom* 93: 509-524.

Goetzmann WN, Jorion P (1993) Testing the predictive power of dividend yields. *J Financ* 48: 663-679.

Hansen PR, Lunde A (2005) A forecast comparison of volatility models: does anything beat a garch (1, 1)? *J Appl Econom* 20: 873-889.

Hillebrand E, Medeiros MC (2016) Asymmetries, breaks, and long-range dependence. *J Bus Econ Stat* 34: 23-41.

Holm S (1979) A simple sequentially rejective multiple test procedure. *Scand J Stat* 6: 65-70.

Hwang E, Shin D (2014) Infinite-order, long-memory heterogeneous autoregressive models. *Comput Stat Data Anal* 76: 339–358.

Kock AB (2016) Consistent and conservative model selection with the adaptive lasso in stationary and nonstationary autoregressions. *Econom Theory* 32: 243-259.

Kock AB, Callot LAF (2015) Oracle inequalities for high dimensional vector autoregressions. *J Econom* 186: 325-344.

Kothari SP, Shanken J (1997) Book-to-market, dividend yield, and expected market returns: A time-series analysis. *J Financ Econ* 44: 169-203.

Leeb H, Pötscher BM (2006) Performance limits for estimators of the risk or distribution of shrinkage-type estimators, and some general lower risk-bound results. *Econom Theory* 22: 69–97.

Leeb H, Pötscher BM (2008) Sparse estimators and the oracle property, or the return of hodges estimator. *J Econom* 142: 201-211.

McAleer M, Medeiros MC (2008) Realized volatility: A review. *Econom Rev* 27: 10-45.

Medeiros MC, Mendes E (2016) l_1-regularization of high-dimensional time-series models with non-gaussian and heteroskedastic innovations. *J Econom* 191: 255-271.

Nelson CR, Kim MJ (1993) Predictable stock returns: The role of small sample bias. *J Financ* 48: 641-661.

Pötscher BM, Schneider U (2009) On the distribution of the adaptive lasso estimator. *J Stat Plan Inference* 139: 2775-2790.

Romano JP, Wolf M (2005) Stepwise multiple testing as formalized data snooping. *Econom* 73: 1237-1282.

Stambaugh RF (1999) Predictive regressions. *J Financ Econ* 54: 375-421.

Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B (Methodol)* 58: 267-288.

Torous W, Valkanov R, Yan S (2004) On predicting stock returns with nearly integrated explanatory variables. *J Bus* 77: 937-966.

Wang SH, Bauwens L, Hsiao C (2013) Forecasting a Long Memory Process subject to Structural Breaks. *J Econom* 177: 171-184.

Zhang L, Mykland PA, Aït-Sahalia Y (2005) A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J Am Stat Associ* 100: 1394-1411.

Zou H (2006) The adaptive lasso and its oracle properties. *J Am Stat Associ* 101: 1418-1429.

Appendix: Adaptive Lasso Statistic

In this appendix, we better clarify the asymptotic properties of the adaptive lasso statistic. To this end, we introduce some results established in Audrino and Camponovo (2015). For a fixed constant $\lambda \geq 0$, consider the adaptive lasso estimator

$$\hat{\beta}_{AL,\lambda} = \arg\min_{\beta} \left\{ \sum_{i=p}^{n} \left(\log RV_{i+1} - \beta_0 - \sum_{j=1}^{p} \beta_j \log RV_{i-j+1}^{(d)} \right)^2 + \lambda \sum_{j=1}^{p} \lambda_j |\beta_j| \right\}.$$

Note that in order to point out that the tuning parameter $\lambda \geq 0$ is fixed, we slightly change the notation introduced in (5), and write $\hat{\beta}_{AL,\lambda}$ instead of $\hat{\beta}_{AL}$. Furthermore, let

$$\Omega_0 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E[\frac{\partial}{\partial\theta}\rho(X_i,\theta)\frac{\partial}{\partial\theta}\rho(X_j,\theta)'],$$

and consider also

$$D_0 = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} E[\frac{\partial^2}{\partial\theta\partial\theta'}\rho(X_t,\theta)],$$

where

$$X_t = (\log RV_{t+1}, \ldots, \log RV_{t-p+1})',$$

$$\theta = (\beta_0, \ldots, \beta_p),$$

$$\rho(X_t,\theta) = \left(\log RV_{i+1} - \beta_0 - \sum_{j=1}^{p} \beta_j \log RV_{i-j+1}^{(d)} \right)^2.$$

Under Assumption 8.1 in Audrino and Camponovo (2015), we can show that

$$\sqrt{n}(\hat{\beta}_{AL,\lambda} - \beta) \to_d \arg\min(R(u)), \tag{8}$$

where

$$R(u) = u'\omega_0 + \frac{u'D_0u}{2} + \sum_{i=1}^{p} \lambda_{0,i}|u_i|,$$

$\omega_0 = (\omega_{0,1}, \ldots, \omega_{0,p})' \sim N(0, \Omega_0)$, for $i = 1, \ldots, p$,

$$\lambda_{0,i} = \begin{cases} 0, & \beta_i \neq 0, \\ \lambda / |\sum_{t=1}^{p} d_{i,t}^I \omega_{0,t}|, & \beta_i = 0, \end{cases}$$

and $d_{i,t}^I$ is the (i,t)-term of the matrix D_0^{-1}. The limit distribution in (8) provides the key result for proving the validity of inference with the adaptive lasso statistic. Indeed, suppose that we are interested in testing the null hypothesis $H_{0,i} : \beta_i = 0$, and consider the adaptive lasso test statistic $T_{\lambda,i} = \sqrt{n}|\hat{\beta}_{AL,\lambda,i}|$. Let $c_{\lambda,i,1-\alpha}$ denote the $1-\alpha$ quantile of the limit distribution of $T_{\lambda,i}$, where $\lambda \geq 0$ is fixed. For instance, when $\lambda = 0$, then $c_{0,i,1-\alpha}$ is simply the $1-\alpha$ quantile of the random variable $|S|$, where $S \sim N(0, v_i)$, and v_i is the i-th diagonal term of $V_0 = D_0^{-1}\Omega_0 D_0^{-1}$. Then, using the limit distribution in (8) we can easily verify that for all $\lambda \geq 0$,

$$c_{0,i,1-\alpha} \geq c_{\lambda,i,1-\alpha},$$

i.e., the $1-\alpha$ quantile $c_{\lambda,i,1-\alpha}$ is maximized at $\lambda = 0$. It turns out that under the null hypothesis $H_{0,i} : \beta_i = 0$,

$$\lim_{n\to\infty} \sup_{0\leq\lambda<\infty} P_{H_{0,i}}(T_{\lambda,i} > z_{i,1-\alpha}) \leq \alpha. \tag{9}$$

Using this result, (conservative) inference based on the adaptive lasso is straightforward. Indeed, to test the null hypothesis $H_{0,i}$ at the significance level $\alpha \in (0,1)$ we can simply use the adaptive lasso test statistic $T_{\lambda,i}$ with the normal critical value $z_{i,1-\alpha}$. As established in (9), the size of this test asymptotically does not exceed the significance level α uniformly over the selection of the tuning parameter λ.

AIMS Press

© 2017 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)