Universität St.Gallen

Lassoing the HAR model: A Model Selection Perspective on Realized Volatility Dynamics

Francesco Audrino and Simon D. Knaus

November 2012 Paper no. 2012-24

School of Economics and Political Science,
Department of Economics

University of St. Gallen

Editor:	Martina Flockerzi University of St. Gallen School of Economics and Political Science Department of Economics Bodanstrasse 8 CH-9000 St. Gallen Phone +41 71 224 23 25 Fax +41 71 224 31 35 Email seps@unisg.ch
Publisher:	School of Economics and Political Science Department of Economics University of St. Gallen Bodanstrasse 8 CH-9000 St. Gallen Phone +41 71 224 23 25
Electronic Publication:	Fax +41 71 224 31 35 http://www.seps.unisg.ch

Lassoing the HAR model: A Model Selection Perspective on Realized Volatility Dynamics[1]

Francesco Audrino and Simon D. Knaus

Author's address:

Prof. Dr. Francesco Audrino
University of St. Gallen
Bodanstrasse 6
CH-9000 St. Gallen
Phone +41 71 224 24 31
Fax +41 71 224 28 94
Email francesco.audrino@unisg.ch

Simon Knaus
University of St. Gallen
Bodanstrasse 6
CH-9000 St. Gallen
Phone +41 71 224 24 75
Fax +41 71 224 28 94
Email simon.knaus@unisg.ch

[1] We thank Matthias Fengler, Daniel Buncic, Pirmin Meier and the participants of 20th International Conference on Computational Statistics (COMPSTAT 2012) in Limassol, Cyprus, for helpful comments.

Abstract

Realized volatility computed from high-frequency data is an important measure for many applications in finance. However, its dynamics are not well understood to date. Recent notable advances that perform well include the heterogeneous autoregressive (HAR) model which is economically interpretable and but still easy to estimate. It also features good out-of-sample performance and has been extremely well received by the research community.

We present a data driven approach based on the absolute shrinkage and selection operator (lasso) which should identify the aforementioned model. We prove that the lasso indeed recovers the HAR model asymptotically if it is the true model, and we present Monte Carlo evidence in finite sample. The HAR model is not recovered by the lasso on real data. This, together with an empirical out-of-sample analysis that shows equal performance of the HAR model and the lasso approach, leads to the conclusion that the HAR model may not be the true model but it captures a linear footprint of the true volatility dynamics.

Keywords

Realized Volatility, Heterogeneous Autoregressive Model, Lasso, Model Selection

JEL Classification

C58, C63, C49

1 Introduction

Volatility of financial assets is of great importance to many applications in finance. Reliable estimates and forecasts are key for risk management and asset allocation. As opposed to return series, financial volatility is predictable and has received great attention in the financial econometrics research community. The seminal paper of Bollerslev (1986) introducing the generalized autoregressive conditional heteroscedasticity (GARCH) model for conditional volatility has thus sparked an even greater interest in volatility modeling. The GARCH model has become extremely popular and despite various extensions and modifications the basic GARCH(1,1) fares well as a prediction device for conditional volatility in an out-of-sample forecast comparison (Hansen & Lunde 2005). While Bollerslev's (1986) GARCH model is able to capture stylized facts of volatility series (e.g., volatility clustering), its estimation still relies on daily observations and thus potentially discards intraday information. The advent of high-frequency data (with frequencies as high as tick-by-tick) has ignited a new line of research pioneered by Andersen, Bollerslev, Diebold & Labys (2001) and Barndorff-Nielsen & Shephard (2002) among others. The results of their work has rendered the thus far unobservable daily volatility observable by means of asymptotic arguments:

Suppose that an asset's log price obeys the dynamics $\mathrm{d}X_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t$ where W_t is a Brownian motion, σ_t the instantaneous volatility and μ_t the instantaneous drift term. One can then show that $\mathrm{plim}_{\delta\to 0} \sum_{t_i} (X_{t_{i+1}} - X_{t_i})^2 = \int_0^T \sigma_t^2 \mathrm{d}t$ where $\delta = \sup\{t_{i+1} - t_i\}$, i.e., the sum of squared returns converges to the integrated variance (over a day) as the sampling frequency increases.[1] An estimator of $\int_0^T \sigma_s^2 \mathrm{d}s$ is thus given by $\sum_{i=1}^N (X_{t_{i+1}} - X_{t_i})^2$ where $t_1, \ldots, t_N$ is an appropriate sampling frequency and is denoted RV_t, where t refers to the day. RV_t is called *realized variance*, and its squareroot $\sqrt{\mathrm{RV}_t}$ is referred to as *realized volatility*. An overview of variants of the aforementioned estimator and their corresponding assumptions is collected in McAleer & Medeiros's (2008) review on realized volatility.

Since the goal of this work is to investigate the dynamics of the realized variance and *not* the estimation itself we can thus – with daily realized variance at hand – approach the problem of

[1] It is known that this naive estimator of $\int_0^T \sigma_t^2 \mathrm{d}t$ is biased under e.g., microstructure noise (the observable return process $Y_{t_i} = X_{t_i} + \varepsilon_{t_i}$ is contaminated with noise) or if the log price process is a jump-diffusion ($\mathrm{d}X_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t + \mathrm{d}J_t$ where J_t is a finite activity jump process).

modeling realized variance.

It has been observed that the time series $\{\mathrm{RV}_t\}_{1\leq t\leq T}$ exhibits some distinct features such as a near log-normal unconditional distribution as well as a slowly decaying autocorrelation function which is often termed "long memory": These findings appear to be robust across different asset classes and evidence has been reported for exchange rates (Andersen, Bollerslev, Diebold & Labys 2001), index futures (Areal & Taylor 2002, Thomakos & Wang 2003), as well as for individual stocks (Andersen, Bollerslev, Diebold & Ebens 2001).

To address these characteristics of the realized variance time series, different approaches have been put forward, most prominently fractionally integrated ARMA models (ARFIMA) and the heterogeneous autoregressive (HAR) model for realized volatility introduced by Corsi (2009). The HAR model not only allows for an economic interpretation of the proposed dynamics, but also allows for an easy estimation and is thus highly appreciated and widely used within the research community.

The contribution of this paper is to shed more light on the underlying dynamics as advocated by Corsi's (2009) HAR model which in essence claims tomorrow's realized variance to be a sum of daily, weekly, and monthly averages of realized variances that can each be attributed to specific investment behaviors. The question we are aiming to answer relates to how much these frequencies (daily, weekly, monthly) are really inherent to the data and if we can identify them from a model selection perspective.

Model selection plays a crucial role in determining a model for forecasting. Oftentimes model selection can be extremely costly from a computational perspective and may already become infeasible within the class of linear models (an exhaustive search over p lags already requires 2^p comparisons and thus grows exponentially). An important contribution in terms of model selection within the class of linear models was made in Tibshirani (1996) where the Least Absolute Shrinkage and Selection Operator (lasso) was introduced. The lasso, a shrunk regression, performs shrinkage and selection at a time and is yet computationally affordable. Although originally the lasso was mostly noticed by the computational statistics community, researchers in econometrics are increasingly using it. Most recently, conditions under which the lasso gives consistent results have also been established in time series econometrics (Nardi & Rinaldo 2011), and applications of the lasso are also found in Park & Sakaori (2012).

Despite the great popularity and appreciation of the HAR model there has been little work investigating the validity of the structure as proposed by the HAR model. Although most work is done in the direction of extending the HAR model (see the recent review of Corsi, Audrino & Renò (2012)) there is a notable exception: Craioveanu & Hillebrand (2010) investigate the structure of the HAR model and find no benefit in allowing for a more flexible structure of lag selection. However, their result is based on an exhaustive search over HAR-like models but varying aggregation frequencies.

It is along these lines that this paper adds to the literature. We present a methodologically sound way of recovering the HAR model. We show that under the assumption that HAR model is the true model, we can apply the lasso and should recover the structure as implied by the HAR model. To this end we investigate how far Nardi & Rinaldo's (2011) can be extended for the special case of the HAR model. Moreover, we investigate if the lasso can be used for forecasting realized variance from a purely statistical point of view as well as measuring outperformance from a more economically relevant point of view via a risk management application. We find no substantial superiority of either the HAR model or the lasso when it comes to out-of-sample forecasting.

In summary, we have reason to believe that the HAR model might not be the true model. However, it captures a linear footprint of the true underlying variance dynamics which appear to change over time, thus casting some doubt on the appropriateness of the HAR as a global model for realized variance.

The rest of the paper is structured as follows: Section 1 introduces the HAR model in more detail, relates it to the autoregressive class of time series models and shows how the lasso can be used in this context. Section 2 features an empirical application of the proposed model selection approach, a Monte CFarlo study, as well as an out-of-sample comparison of the HAR versus the lasso. Section 3 discusses the results and further research and then concludes.

2 Theoretical Foundation

2.1 The HAR Model

The HAR model as introduced in Corsi (2009) enjoys great popularity: It allows for an economic interpretation, has good forecasting performance, and is still easy to estimate. There are numerous variants and modifications of the HAR model (Corsi et al. 2012), however we restrict our attention to the original model to keep a clear focus on the actual volatility dynamics. We thus intentionally ignore other transient effects (such as the leverage effect) that may be embedded in a HAR framework as well.

Let for this purpose $\mathrm{RV}_t^{(d)}$ be an estimate of daily realized variance. Then, the HAR model postulates that

$$\log \mathrm{RV}_{t+1}^{(d)} = c + \beta^{(d)} \log \mathrm{RV}_t^{(d)} + \beta^{(w)} \log \mathrm{RV}_t^{(w)} + \beta^{(m)} \log \mathrm{RV}_t^{(m)} + \omega_{t+1}, \tag{1}$$

where (with a slight abuse of notation) $\log \mathrm{RV}_t^{(w)} = \frac{1}{5}\sum_{i=1}^{5} \log \mathrm{RV}_{t-i+1}^{(d)}$ and $\log \mathrm{RV}_t^{(m)} = \frac{1}{22}\sum_{i=1}^{22} \log \mathrm{RV}_{t-i+1}^{(d)}$ are the weekly and monthly averages of daily log realized variances, and ω_{t+1} is an innovation. Once these average log-variances are known, the model can be consistently estimated by traditional least squares to obtain estimates for c, $\beta^{(d)}$, $\beta^{(w)}$, and $\beta^{(m)}$.

In other words, the conditional expectation of tomorrow's log-realized variance is the weighted sum (plus an intercept) of daily, weekly, and monthly log-realized volatilities.[2] For the remainder of the paper we assume the HAR model to be causal as well as $\beta^{(d)}$, $\beta^{(w)}$, $\beta^{(m)}$ to be positive. These assumptions are by no means restrictive: First they comply with the view put forward in the original work as outlined below, second, if estimating the HAR on empirical data, the coefficients are always found to be positive.

The different aggregation frequency can then be seen as a heterogeneous agent model where heterogeneity is induced by the different time horizons and can be casted into an information cascade view. Hence, the weighted average perspective appears reasonable and positiveness of the coefficients follows.

Clearly, the HAR model is simply a constrained AR(22) model, as it has already been noted

[2] We comment further on the use of log-realized volatilities in Section 3.1.

by Corsi (2009), i.e., we can write

$$\log \mathrm{RV}_{t+1}^{(d)} = \phi^{\mathrm{HAR}} + \sum_{i=1}^{22} \phi_i^{\mathrm{HAR}} \log \mathrm{RV}_{t-i+1}^{(d)} + \omega_{t+1} \tag{2}$$

where the restrictions as imposed by (1) require

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

A direct specification test is obviously testing the restrictions as collected by (3). Given the high number of restrictions a rejection of these is not surprising. However, in the original work Corsi argues that this can well be attributed to specific properties of the time series. However, there is already some preliminary indication that indeed the HAR model may fail to fully capture the effects present in the data.

2.2 The lasso as model selection device

The lasso was introduced in Tibshirani (1996) and is frequently used in the field of computational statistics and machine learning. In recent years, the lasso in general as well as the lasso as model selection device has also been found in Econometrics (Kock 2012, Leeb & Pötscher 2005). The lasso is computationally very efficient and renders model selection with a high number of predictors feasible. As opposed to the 2^p comparisons that are required in an exhaustive search over p predictors, the lasso employs a highly efficient algorithm which provides estimates and model selection jointly (Friedman, Hastie & Tibshirani 2010) at affordable computational costs.

The lasso as originally introduced by Tibshirani covered the cross-sectional case: Let $\mathbf{x_i} = (x_{i1}, \ldots, x_{ip})'$ be predictor variables and y_i responses. Under the assumption that the predictors are standardized, the lasso estimator of the model

$$y_i = \alpha + \phi' \cdot x_i + \epsilon_i \tag{4}$$

is obtained as

$$(\hat{\alpha}^{\text{lasso}}, \hat{\phi}^{\text{lasso}}) = \arg\min_{\alpha,\phi} \left\{ \sum_{i=1}^{n} \left(y_i - \alpha - \sum_{j=1}^{p} \phi_j x_{ij} \right)^2 \right\} \quad \text{subject to} \sum_{j=1}^{p} |\phi_j| \leq t \tag{5}$$

where t is a tuning parameter. Since $\hat{\alpha}$ is independent of t it will always be equal to $\bar{y}$ and it is thus generally assumed that $\bar{y} = 0$ and α is dropped from the minimization. It can be seen (Tibshirani 1996) that (5) is equivalent to the Lagrangian form given as

$$\hat{\phi}^{\text{lasso}} = \arg\min_{\phi} \left\{ \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \phi_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\phi_j| \right\} \tag{6}$$

with a one-to-one correspondence between λ in (6) and t in (5). The powerful feature of the lasso is now induced by the L^1-norm of the penalty. The lasso solution will be sparse, since some ϕ_js will be set exactly to zero (as opposed to for instance ridge regularization in Hastie, Tibshirani & Friedman (2009) where sparsity of the solution is lost due to the L^2-geometry of ridge).

A question of utmost importance is how reliable is the lasso in the sense that it sets the true zero coefficients to zero. Typically, this is what is captured by *model selection consistency*. The following definition adopts the view of Nardi & Rinaldo (2011). For an overview and weaker form of this, the reader is referred to Bühlmann & Van De Geer (2011).

Definition 1. *Let* $y_i = \phi' \cdot x_i + \epsilon_i$ *with* $\phi^0 = [\phi_1^0, \ldots, \phi_p^0]'$, $\text{sgn} : \mathbb{R} \to \{-1, 0, 1\}$ *and define* $\text{sgn}(\phi) = (\text{sgn}(\phi_1), \ldots, \text{sgn}(\phi_p))'$. *Then an estimator* $\hat{\phi}_n$ *is said to be* model selection consistent *if*

$$\mathrm{P}(\text{sgn}(\hat{\phi}_n) = \text{sgn}(\phi^0)) \to 1 \textit{ for } n \to \infty. \tag{7}$$

The above model selection consistency definition meets our requirement that if there is an estimator producing $\hat{\phi}_n$ which is *model selection consistent* it will eventually only retain the true non-zero coefficients $\operatorname{supp} \phi^0$.

An extension of the lasso as well as proof for which conditions the lasso is model selection consistent is given in Zou (2006). Zou introduces the adaptive lasso which allows for a more flexible penalization, i.e.,

$$\hat{\beta} = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \phi_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \lambda_j |\phi_j| \right\} \quad (8)$$

where λ_j are adaptive weights. It can be shown (Zou 2006, Bühlmann & Van De Geer 2011) that in fact the adaptive lasso relaxes the assumptions for the model selection consistency of the lasso.

An important extension of this strand of literature has been made by Nardi & Rinaldo (2011): Nardi & Rinaldo show that properties already well-established in the cross sectional case carry over to the time series case of an AR(p) process.[3] More precisely, they establish that under some assumptions, a version of the adaptive lasso is model selection consistent. Suppose that X_t is a causal Gaussian AR(p) process, i.e.,

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \epsilon_t$$

where ϵ_t is i.i. $\mathcal{N}(0, \sigma^2)$-distributed. Define $S = \{j, \phi_j \neq 0\} \subset \{1, \ldots, p\}$ the active set, $S^c = \{1, \ldots, p\} \setminus S$ the non-active set, and $\Gamma_{XY} = \text{Cov}(X, Y)$ the covariance matrix of a vector X and Y. Consequently, Γ_{SS} is the square covariance matrix of the active predictors and Γ_{S^cS} is the covariance matrix of the predictors in the non-active set (given as $\{X_{t-j},\ j \in S^c\}$) with the predictors in the active set (given as $\{X_{t-j},\ j \in S\}$). They then proceed and prove the following theorem (Nardi & Rinaldo 2011, Theorem 3.1):

Theorem 1. *Consider the AR(p) settings described above. Assume that*

(i) there exists a finite positive constant $C_{\max}$ *such that* $\|\Gamma_{SS}^{-1}\|_\infty \leq C_{\max}$*;*

(ii) there exists a $\delta \in (0, 1]$ *such that* $\|\Gamma_{S^cS}\Gamma_{SS}^{-1}\|_\infty \leq 1 - \delta$*.*

Further assume that the asymptotic properties for λ_n *and* $\lambda_{n,j}$ *as given in Nardi & Rinaldo (2011, Theorem 3.1) hold.*

Then, the lasso estimator is model selection consistent in the sense of Definition 1.

[3]Note that Definition 1 is by no means limited to the cross-sectional case and translates directly to the time series regression variant.

Condition (ii) of the above theorem is found throughout the model consistency literature for the lasso. Typically this condition is called the *irrepresentable condition* as introduced in Zhao & Yu (2006). Nardi & Rinaldo show that a causal Gaussian process satisfies the assumptions of Theorem 1 and the lasso is thus *model selection consistent* for this class of models.

2.3 Lassoing the HAR model

Theorem 1 states that the lasso is indeed model selection consistent for causal AR(p) processes with Gaussian innovations. If we assume that ϵ_t in (1) is Gaussian we can readily use the lasso to try to recover the HAR model embedded in an AR(p) process with $p > 22$. The lasso should then detect[4] $S = \{1, 2, \ldots, 22\}$ and $S^c = \{23, 24, \ldots, p\}$ since any other lagged value should be irrelevant if the HAR model is the true data generating process (DGP).

The assumption of Gaussianity of the error may appear strong at first sight. However, the HAR model is usually estimated using quasi-likelihood which in turn also assumes Gaussianity. An even stronger argument is given below and proved in the appendix. Under the assumption that the HAR model is the true DGP, we precisely know the dynamics and can prove (ii) of Theorem 1 directly without relying on Gaussianity. This can then be used in Zhao & Yu's (2006) result which relaxes the assumptions of Gaussianity of the innovations. The relaxation on the distribution of the error term comes at the price of keeping S and S^c fixed; the lasso literature generally differentiates between a $p = |S|$ growing with n or p fixed. Theorem 1 above addresses the case where p is allowed to grow, our contribution below however requires p to be fix:

Theorem 2. *Under the assumptions that the DGP is as given in* (1) *is causal and the innovation has a finite fourth moment, S^c is held fixed, then lasso is model selection consistent in the sense of Definition 1.*

The complete argument and proof is given in Appendix A.

3 Empirical Application

In this section we illustrate our approach of identifying the HAR model via the lasso using nine assets traded on the New York Stock Exchange. For each of these stocks we compute a realized

[4] In the sense of setting the non-active coefficients to zero.

variance measure using Zhang, Mykland & Aït-Sahalia's (2005) two-time scales estimator (using a frequency of 10 minutes) to obtain a series of daily realized variance measures. These measures are then used to estimate the HAR model in-sample and contrast it with estimates as obtained by the lasso procedure described in Section 2. We also compare the lasso's forecasting performance to the performance of the HAR out-of-sample. To rule out any doubt that these findings are dependent on a specific realized variance estimator we also report a summary of results using Andersen, Dobrev & Schaumburg's (2009) MedRV estimator in Appendix C. The key descriptive properties of the data are summarized in Fig. 1 and Tab. 1.

Note that we obviously only forecast one day ahead realized variance since our argument is based on the original specification of the HAR model. One could of course address the question whether the lasso is also well suited to forecast realized variance at longer horizons (weekly, monthly); this however would be a purely empirical exercise and is beyond the scope of this paper.

3.1 Data Description

We use intraday data of Alcoa, Inc. (AA), Citigroup, Inc. (C), Hasbro Inc. (HAS), The Home Depot, Inc. (HDI), Intel Corporation (INTC), Microsoft Corporation (MSFT), Nike Inc. (NKE), Pfizer Inc. (PFE), and Exxon Mobil Corporation (XOM) from Jan 2, 2001 to Nov 15, 2010. These intraday data are then used to compute an estimator of daily realized variance using Zhang et al.'s (2005) two-time scales estimator. In total we have 2483 observations of realized variance measures.

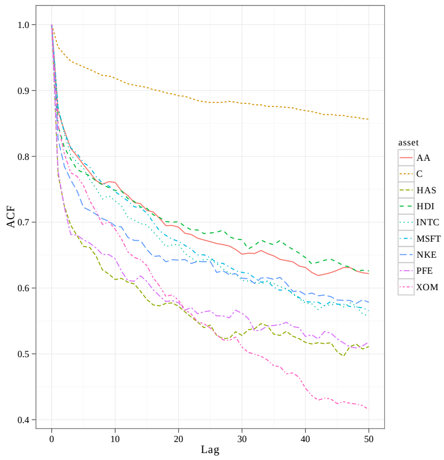

Figure 1 (a)

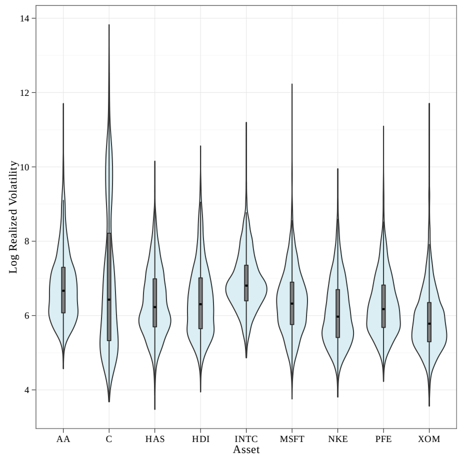

Figure 1 (b)

Figure 1: Panel (a) shows the autocorrelation function for the 9 $\log \mathrm{RV}_t$ series. Panel (b) shows a violin plot (Hintze & Nelson 1998) of the unconditional $\log \mathrm{RV}_t$

Table 1: Descriptive Statistics of $\log \mathrm{RV}_t$ series

	AA	C	HAS	HDI	INTC	MSFT	NKE	PFE	XOM
Mean	6.80	6.96	6.36	6.42	6.90	6.37	6.10	6.30	5.90
SD	0.95	1.98	0.93	0.98	0.78	0.89	0.90	0.85	0.89
Kurtosis	4.06	2.61	3.06	3.32	3.82	4.40	3.06	4.18	6.15
Skewness	0.92	0.78	0.46	0.69	0.58	0.58	0.58	0.83	1.19
Median	6.67	6.43	6.23	6.30	6.81	6.33	5.97	6.18	5.78
25%-quantile	6.07	5.33	5.70	5.65	6.40	5.76	5.41	5.68	5.30
75%-quantile	7.30	8.21	6.99	7.01	7.35	6.90	6.70	6.82	6.35

Although using the log to transform the realized variance is standard in the literature, we briefly comment explicitly on this in Appendix B for the HAR model. In what follows we always assume the use of log realized variance when speaking of realized variance unless otherwise stated.

Consistent with the existing literature we witness slowly decaying autocorrelation functions in Fig. 1 (a) for all assets. This is most pronounced for Citigroup, Inc. The same stock also exhibits particularities in the unconditional distribution of $\log \mathrm{RV}_t$ as can be seen from Tab. 1

and Fig. 1 (b): While all other stocks show excess kurtosis, Citigroup Inc. only has a kurtosis of 2.61. We suspect the market turmoil of the financial crisis to be the root of this abnormal picture. Following this train of thought, we also report the actual returns in Fig. 10 in the appendix where an extremely high excess kurtosis for the log returns of Citigroup Inc. can be observed.

3.2 In-sample Evaluation

To address the question whether the HAR model is identified by the lasso procedure we define $S^c = \{x_{t-23}, \ldots, x_{t-100}\}$.[5] Since λ in (6) is a tuning parameter and the results of Theorem 1 only hold asymptotically we proceed as suggested in the literature (Nardi & Rinaldo 2011, Section 4.1.) and choose $\lambda_j = 1$ for all j and $\lambda = \sqrt{\frac{\log n \log p}{n}}$ and can thus expect $\hat{S}$ as obtained by $\hat{\phi}^{\text{lasso}}$ to be sparse in $\{1, \ldots, 100\}$.

Two important points should be noted here: First, the lasso does not recover all of the coefficients implied to be non-zero by the HAR as can be inferred from Tab. 2. Although near lags are recovered for most assets, lags beyond x_{t-6} rarely get selected by the lasso. Note at this point that a comparison of coefficients in magnitude of the lasso estimates to the HAR estimates cannot be made since the lasso, as a penalized estimator, is biased. Second, sometimes lags far beyond x_{t-22} are selected in the active set as can be seen in Fig. 2. Clearly, these lags are zero under the assumption that the HAR model is true.

At this stage it is alaredy apparent that the lasso does *not* fully recover the HAR model, i.e. $\hat{S} \neq \{1, \ldots, 22\}$. To provide further evidence supporting this statement, we conduct analyses which attempt to answer the following two questions: 1. How reliable is the lasso as a model selection device in this specific finite sample setting? 2. How stable are these regressors over time? A thorough answer to these questions is provided in the two subsequent paragraphs.

3.2.1 Monte Carlo Study

To assess the model selection consistency of the lasso in the case of the HAR model in finite sample we include a Monte Carlo simulation in this section. Since the lasso's model selection results depend on the signal-to-noise ratio (Bühlmann & Van De Geer 2011), it is important

[5] The choice of S running up to 100 is arbitrary. However, the results are not sensitive to the choice of the maximal lag, as for instance the results remain almost identical for a maximal lag of 50

Table 2: HAR estimates versus lasso estimates

	AA		C		HAS		HDI		INTC		MSFT		NKE		PFE		XOM	
Lag	HAR	lasso	HAR	lasso	HAR	lasso	HAR	lasso	HAR	lasso	HAR	lasso	HAR	lasso	HAR	lasso	HAR	lasso
γ_1	0.470	0.418	0.573	0.530	0.403	0.358	0.413	0.370	0.535	0.483	0.488	0.444	0.456	0.408	0.417	0.376	0.497	0.455
γ_2	0.084	0.172	0.073	0.169	0.084	0.121	0.086	0.137	0.074	0.118	0.086	0.125	0.073	0.116	0.073	0.112	0.094	0.130
γ_3	0.084	0.039	0.073	0.001	0.084	0.072	0.086	0.077	0.074	0.003	0.086	0.042	0.073	0.076	0.073	—	0.094	0.033
γ_4	0.084	0.061	0.073	0.065	0.084	0.040	0.086	0.019	0.074	0.062	0.086	0.087	0.073	0.049	0.073	0.067	0.094	0.096
γ_5	0.084	0.029	0.073	0.044	0.084	0.022	0.086	0.050	0.074	0.049	0.086	0.014	0.073	—	0.073	0.039	0.094	0.052
γ_6	0.010	0.010	0.008	0.024	0.012	0.053	0.012	0.052	0.008	—	0.008	0.051	0.013	0.027	0.014	0.040	0.005	—
γ_7	0.010	—	0.008	—	0.012	0.032	0.012	0.012	0.008	—	0.008	—	0.013	0.031	0.014	0.025	0.005	—
γ_8	0.010	—	0.008	—	0.012	—	0.012	0.013	0.008	—	0.008	—	0.013	0.005	0.014	0.005	0.005	—
γ_9	0.010	0.043	0.008	0.056	0.012	—	0.012	0.029	0.008	0.042	0.008	0.040	0.013	0.023	0.014	0.036	0.005	0.039
γ_{10}	0.010	0.056	0.008	0.001	0.012	—	0.012	0.013	0.008	—	0.008	0.002	0.013	0.004	0.014	0.027	0.005	—
γ_{11}	0.010	—	0.008	—	0.012	0.021	0.012	0.030	0.008	—	0.008	0.002	0.013	0.025	0.014	—	0.005	—
γ_{12}	0.010	—	0.008	—	0.012	—	0.012	—	0.008	—	0.008	—	0.013	—	0.014	—	0.005	—
γ_{13}	0.010	—	0.008	—	0.012	0.011	0.012	0.003	0.008	—	0.008	—	0.013	—	0.014	—	0.005	—
γ_{14}	0.010	0.005	0.008	0.005	0.012	—	0.012	—	0.008	—	0.008	0.006	0.013	0.008	0.014	0.030	0.005	—
γ_{15}	0.010	—	0.008	—	0.012	—	0.012	—	0.008	—	0.008	—	0.013	—	0.014	—	0.005	—
γ_{16}	0.010	—	0.008	—	0.012	—	0.012	—	0.008	—	0.008	—	0.013	—	0.014	—	0.005	—
γ_{17}	0.010	—	0.008	—	0.012	—	0.012	—	0.008	—	0.008	—	0.013	—	0.014	—	0.005	—
γ_{18}	0.010	—	0.008	—	0.012	—	0.012	—	0.008	—	0.008	—	0.013	—	0.014	—	0.005	—
γ_{19}	0.010	—	0.008	—	0.012	0.007	0.012	—	0.008	—	0.008	—	0.013	—	0.014	—	0.005	—
γ_{20}	0.010	—	0.008	—	0.012	—	0.012	—	0.008	—	0.008	—	0.013	—	0.014	—	0.005	—
γ_{21}	0.010	—	0.008	0.001	0.012	—	0.012	—	0.008	—	0.008	—	0.013	0.008	0.014	—	0.005	—
γ_{22}	0.010	—	0.008	—	0.012	—	0.012	—	0.008	—	0.008	—	0.013	—	0.014	—	0.005	—

This table reports the HAR coefficients (as implied by (3)) and the lasso coefficients. Coefficients set to 0 by the lasso procedure are indicated by dash. We only report the coefficients up to lag x_{t-22}. Lasso coefficient estimates of lags higher than x_{t-22} are reported graphically in Fig. 2

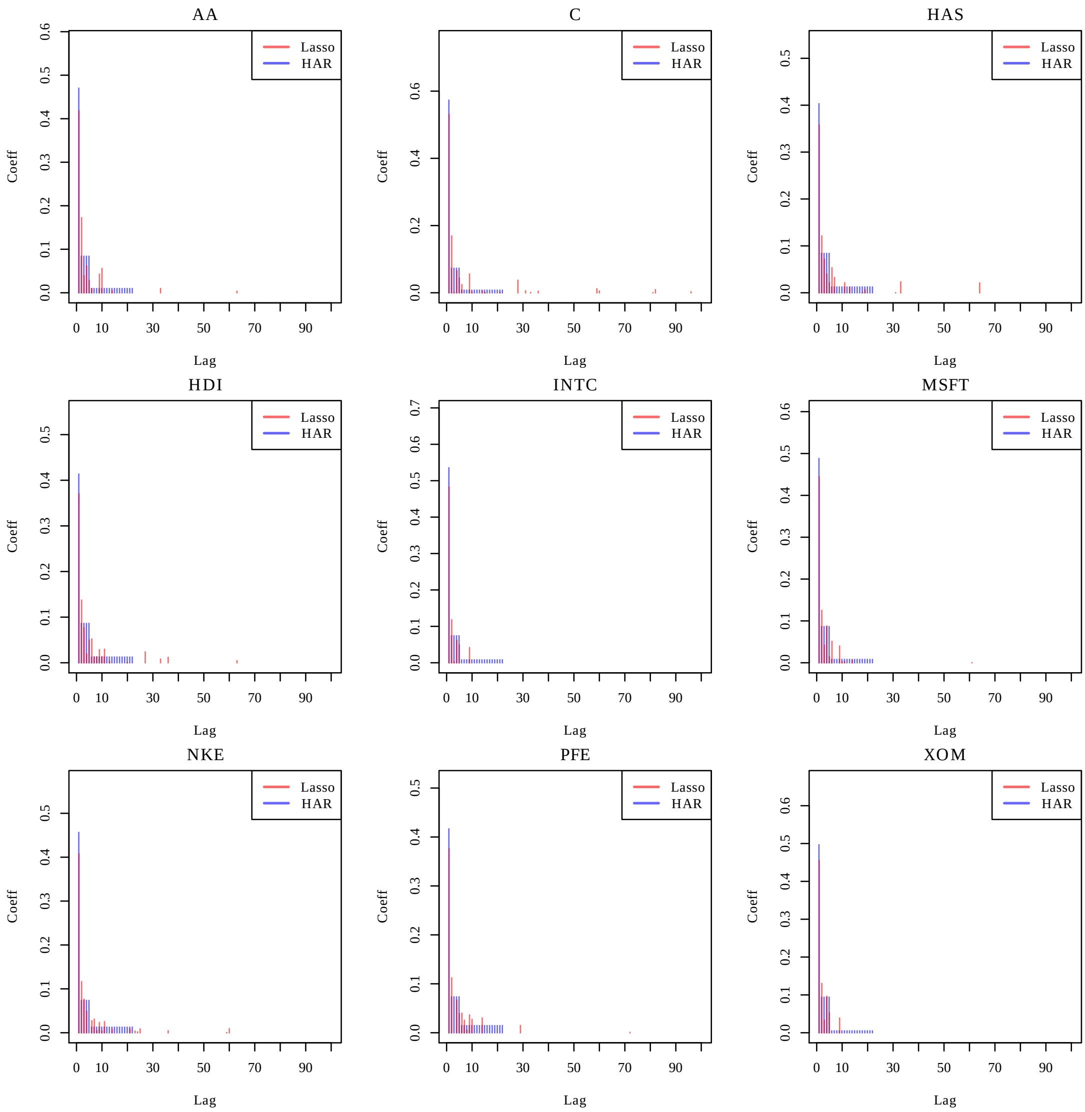

Figure 2: HAR versus lasso coefficients with all predictors

to have a comparable setting to assess the finite sample performance of the lasso as a model selection device. We conducted the Monte Carlo study under the assumption that the HAR model was true, in order to answer the question how effective the lasso *would* be if the HAR model *were* true. To this end, we proceeded as follows in a parametric bootstrap manner:

1. For asset $j = 1, \ldots, 9$ estimate the HAR model on the full sample of 2483 data points, which includes
   (a) Obtain $c, \hat{\beta}^{(d)}, \hat{\beta}^{(m)}, \hat{\beta}^{(w)}$ and compute $\widehat{\mathrm{Var}}(\epsilon_t)$ as well as the derived estimates $\hat{\phi}_1^{(HAR)}, \ldots, \hat{\phi}_{22}^{(HAR)}$ via (3).
   (b) Compute the unconditional mean $\hat{\mu}$ (as $\hat{\gamma}_0/(1 - \sum_{i=1}^{22} \hat{\phi}_i)$) and the unconditional variance $\hat{\sigma}$ (as $\widehat{\mathrm{Var}}(\epsilon_t)/(1 - \sum_{i=1}^{22} \hat{\phi}_i \hat{\gamma}_i)$ where $\hat{\gamma}_i$ is the autocovariance at lag i, see Brockwell & Davis (1986))
2. Resample the HAR model.
   (a) Sample $x_1, \ldots, x_{22}$ from the stationary distribution $\mathcal{N}(\hat{\mu}, \hat{\sigma})$
   (b) Compute $x_{23}, \ldots, x_{2483}$ recursively based on (3).
   (c) Apply the lasso as specified in Section 3.2 and record the lasso estimates

Step 2 is repeated 1,000 times and the results are reported in Tab. 3.The results clearly indicate that the HAR structure is well recovered by the lasso in this synthetic HAR setting. Although small coefficients (the monthly coefficients) are selected less often, the daily and weekly coefficients are almost always estimated to be non-zero and thus considered active. Note at this point that there is indeed some contradiction with what has been reported in Tab. 2: The lasso does not select $\gamma_1, \ldots, \gamma_5$ for all assets and selection of lags beyond 22 is rare.[6]

We thus conclude from this Monte-Carlo application that indeed the lasso does recover the HAR model reasonably well *if* it is the true model, i.e., if we simulate from this DGP.

[6]Based on the percentage of times recovered we may conclude that for instance lag x_{t-15} is non-active across all nine assets (as found in Tab. 2) has a chance of occurring of 6.7% based on the occurrences in Tab. 3.

Table 3: Percentage of HAR coefficients recovered

Lag	AA	C	HAS	HDI	INTC	MSFT	NKE	PFE	XOM
x_{t-1}	100	100	100	100	100	100	100	100	100
x_{t-2}	100	100	100	100	100	100	100	97	100
x_{t-3}	100	100	100	100	100	100	100	98	100
x_{t-4}	100	100	100	100	100	100	100	97	100
x_{t-5}	100	100	100	100	99	99	100	96	100
x_{t-6}	42	43	61	54	18	21	52	23	12
x_{t-7}	37	39	59	53	17	18	50	21	9
x_{t-8}	36	37	61	54	15	16	50	20	6
x_{t-9}	32	32	54	49	9	10	44	15	2
x_{t-10}	34	34	56	50	9	10	45	16	1
x_{t-11}	31	31	54	45	7	9	42	14	1
x_{t-12}	28	30	56	48	7	9	44	15	1
x_{t-13}	27	28	55	47	5	7	43	14	1
x_{t-14}	28	29	58	51	6	7	46	13	0
x_{t-15}	25	24	55	46	4	4	42	10	0
x_{t-16}	25	25	53	44	4	5	40	12	0
x_{t-17}	19	20	51	42	2	2	36	8	0
x_{t-18}	20	20	51	43	2	2	38	8	0
x_{t-19}	16	17	46	36	1	2	31	6	0
x_{t-20}	12	12	42	31	1	1	27	3	0
x_{t-21}	10	12	41	29	0	0	25	3	0
x_{t-22}	8	9	34	25	0	0	20	2	0
x_{t-23}	3	6	13	8	0	0	6	1	0
x_{t-24}	2	4	8	6	0	0	4	0	0
x_{t-25}	1	4	8	5	0	0	3	0	0
x_{t-26}	1	3	5	4	0	0	3	0	0
x_{t-27}	0	2	4	2	0	0	2	0	0
x_{t-28}	0	2	4	3	0	0	1	0	0
x_{t-29}	1	2	4	2	0	0	2	0	0
x_{t-30}	0	1	3	2	0	0	1	0	0
x_{t-31}	0	2	3	2	0	0	1	0	0
x_{t-32}	0	1	3	1	0	0	1	0	0
x_{t-33}	0	1	2	1	0	0	1	0	0
x_{t-34}	0	1	2	1	0	0	0	0	0
x_{t-35}	0	1	1	0	0	0	0	0	0
x_{t-36}	0	1	2	1	0	0	1	0	0
x_{t-37}	0	1	2	1	0	0	1	0	0
x_{t-38}	0	1	2	1	0	0	0	0	0
x_{t-39}	0	1	1	0	0	0	0	0	0
x_{t-40}	0	0	1	0	0	0	0	0	0
x_{t-41}	0	0	1	0	0	0	0	0	0
x_{t-42}	0	0	1	0	0	0	0	0	0
x_{t-43}	0	0	1	0	0	0	0	0	0
x_{t-44}	0	1	1	0	0	0	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{t-100}	0	0	0	0	0	0	0	0	0

Number of times (out 1,000 replications) a lag has been selected (estimated as non-zero) by the lasso in percent. Omitted rows contain zero only.

3.2.2 Rolling Window

To address the question whether all of the observed in-sample selected regressors are constant over time we apply the lasso procedure in a rolling window manner. We stack our data for each asset as follows

$$X = \begin{bmatrix} x_{101} & x_{100} & \dots & x_1 \\ x_{102} & x_{101} & \dots & x_2 \\ \vdots & \vdots & & \vdots \\ x_n & x_{n-1} & \dots & x_{n-100} \end{bmatrix}$$

We then estimate the lasso on the first 1,000 rows of X and roll this window of length 1,000 down to the last row of X. Pursuing this procedure we obtain 1,384 lasso estimates and record them. Fig. 3 contains this analysis for Citigroup, Inc. The abscissa reports the last date of the current window (the first window thus corresponds to the date of x_{1000} which in this case is May 19, 2005 and continues through Nov 15, 2010), the ordinate indicates whether or not a regressor was selected (estimated to be different from zero).

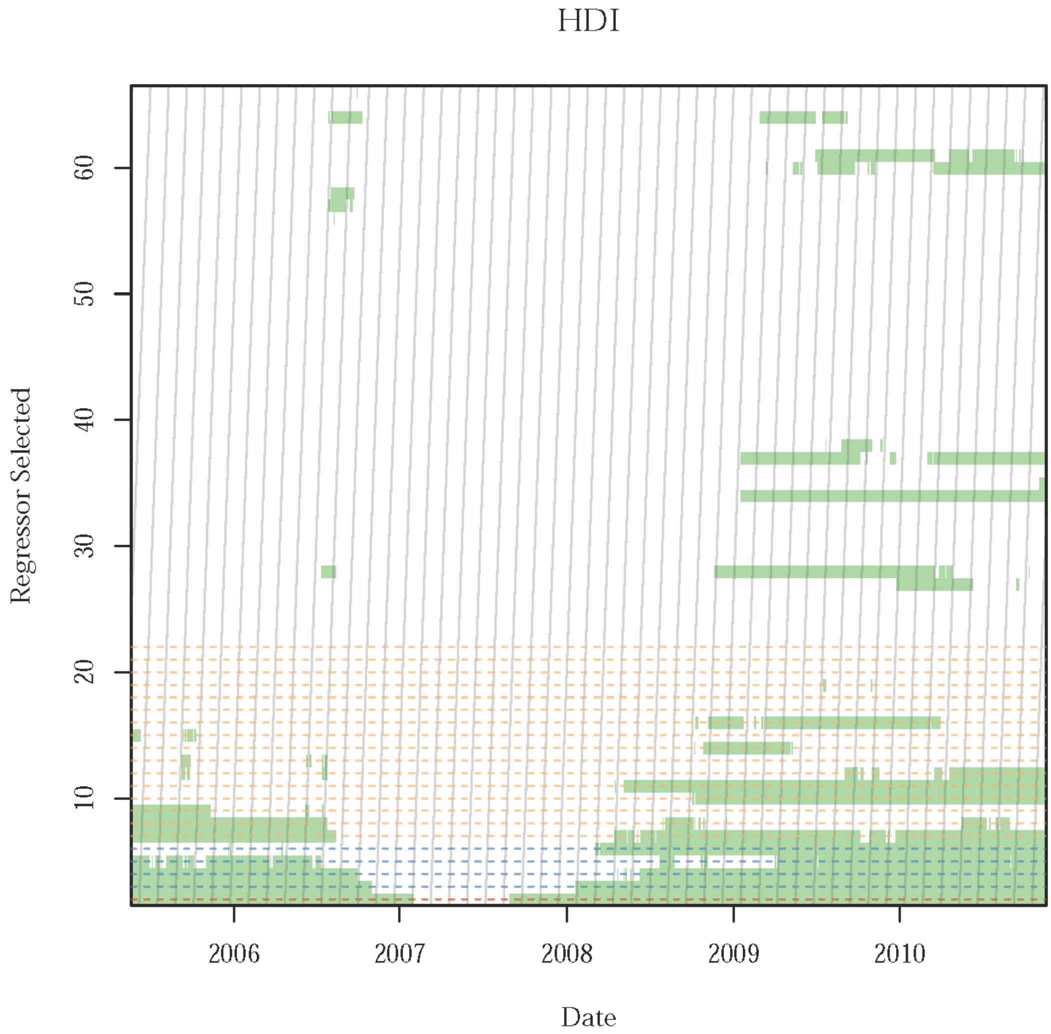

Figure 3: Stability of lasso selected regressors for Home Depot, Inc. Diagonal gray lines have slope 1, i.e., if a regressor moves along these lines then its effect is lagged by one day as the rolling window proceeds by one row (1 day)

Groups of regressors moving along the diagonal lines are likely to be noise (they are one-off events that move through the sampling window). It is also apparent from Fig. 3 that there is a clear break in structure during the financial crisis. The only lag which is selected during the crisis is the x_{t-1} indicating that the variance process prevailing in the data is actually an

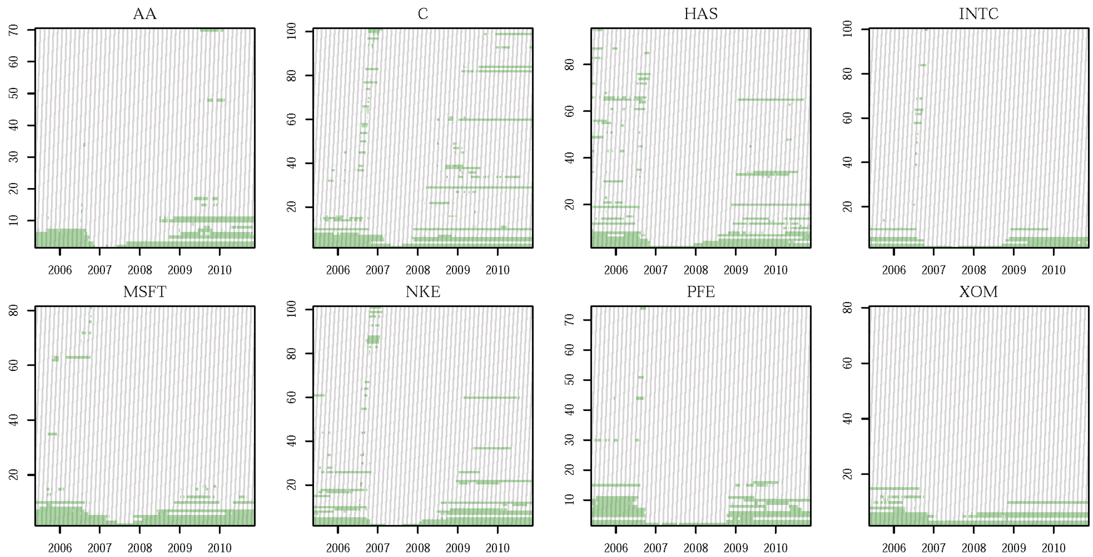

Figure 4: Stability of Lasso selected Regressors for all assets

AR(1)-process.

Fig. 4 draws the same picture for the remaining eight assets. Although there are minor differences among assets we observe a clear pattern of a "dependence breakdown" during the financial crisis. Most assets indeed also have components that can be explained by one-off events, however, we also find for HAS, HDI, and C lags that constantly get selected and remain (beyond the training window length of 1,000 observations). This may be an indication of longer-range dependence that warrants further research. As can clearly be inferred from Fig. 4 the dependence breakdown during the financial crisis is for some assets even more pronounced than it is for Citigroup, Inc. For these, the optimal lag structure as chosen by the lasso, sometimes reduces to a constant (e.g., HDF in Fig. 4). Also, there are assets that exhibit a dependence structure (i.e., by lags beyond x_{t-22}) which is not accounted for by the HAR model.

3.3 Out-of-sample prediction

So far we have only considered the lasso results in-sample. But the HAR has also garnered prais for its for out-of-sample prediction. In a next step we thus compare the HAR's and the lasso's out-of-sample performance. We estimate the HAR model with data up to time t and compute an estimate for $t+1$ which is labeled $\widehat{\log \mathrm{RV}}^{(HAR)}_{t+1|t}$. We do the same for the lasso to obtain $\widehat{\log \mathrm{RV}}^{(lasso)}_{t+1|t}$. We proceed again in a rolling window manner but also vary the training window length (the length on which we estimate the lasso and the HAR model). To render the results comparable we report the out-of-sample prediction for different training window length but the same evaluation window (from May 12, 2009 to Nov 15, 2010 as implied by the longest training window length and resulting in 383 observations) in Tab. 4. To have an objective comparison we also include the random walk in our analysis. Although there is theoretical guidance for choosing λ in (6) we pursue a different approach. The theoretical guidance is optimal in the sense of asymptotic model selection consistency; however, this is not necessarily the best penalty for prediction. Thus, we employ the common approach of estimating the expected prediction error using cross validation.

Cross-validation in the cross sectional case is a statistically sound way of estimating the expected out-of-sample prediction error and thus determining the optimal penalty parameter (Arlot & Celisse 2010, Hastie et al. 2009). Although cross-validation (typically K-fold) is often used in practice to determine the optimal penalty parameter in a penalized regression setting (for instance in Nardi & Rinaldo (2011) and Park & Sakaori (2012)) we adopt the view of Bergmeir & Benítez (2012) and use blocked cross validation[7] to account for the time series nature of the data. When comparing the estimates of $\hat{\lambda}_{\mathrm{opt}}$ obtained by using the regular K-fold cross validation ($\hat{\lambda}^{(R)}_{\mathrm{opt}}$) to the estimates obtained used a K-fold blocked cross-validation ($\hat{\lambda}^{(B)}_{\mathrm{opt}}$), we observed that $\hat{\lambda}^{(R)}_{\mathrm{opt}} < \hat{\lambda}^{(B)}_{\mathrm{opt}}$. From a conceptual point of view, this observation is in accordance with the result that for kernel regression the bandwidth is smaller for positively correlated errors when compared to uncorrelated errors (Hart & Wehrly 1986). Even if kernel regression and the lasso may at first appear as different approaches they can be related, exploiting the linearity of both approaches, by looking at the trace of their smoother matrix (the generalized cross-validation, GCV) which again is an estimate of the prediction error (Hastie et al. 2009).

Summarizing, we use blocked cross validation for both, empirically and theoretically founded reasons, to obtain an optimal λ in our out-of-sample procedure. We use 10 blocks to find an estimate of the optimal λ.

[7]Instead of building K blocks by randomly assigning any number in $\{1, \dots, K\}$ to each observation and collecting the observations having the same number we use blocks with contiguous observations, such that the blocks are $\{1, \dots, K\}, \{K+1, \dots, 2K\}, \dots, \{(\lfloor n/K \rfloor - 1)K + 1, \dots, n\}$

Table 4: Out-of-sample comparison

	200			400			1,000			2,000		
Asset	RW	HAR	lasso	RW	HAR	lasso	RW	HAR	lasso	RW	HAR	lasso
AA	0.160	0.129	0.142	0.160	0.126	0.126	0.160	0.125	0.123	0.160	0.124	0.121
C	0.132	0.115	0.127	0.132	0.115	0.120	0.132	0.115	0.119	0.132	0.116	0.116
HAS	0.240	0.201	0.219	0.240	0.197	0.204	0.240	0.193	0.197	0.240	0.197	0.200
HDI	0.231	0.184	0.207	0.231	0.181	0.186	0.231	0.179	0.181	0.231	0.179	0.178
INTC	0.113	0.094	0.100	0.113	0.091	0.091	0.113	0.089	0.088	0.113	0.089	0.087
MSFT	0.153	0.128	0.137	0.153	0.125	0.127	0.153	0.123	0.124	0.153	0.123	0.121
NKE	0.176	0.144	0.158	0.176	0.142	0.146	0.176	0.139	0.140	0.176	0.138	0.140
PFE	0.130	0.107	0.112	0.130	0.104	0.105	0.130	0.102	0.101	0.130	0.102	0.099
XOM	0.221	0.182	0.192	0.221	0.179	0.179	0.221	0.178	0.175	0.221	0.176	0.174

MSPE for all nine assets across training window length of 200, 400, 1,000, and 2,000 observations (rolling window). In addition to the lasso and the HAR the random walk (RW) is included.

We measure the out-of-sample performance using the mean squared prediction error (MSPE) which is computed as MSPE $= \frac{1}{n}\sum_{t=1}^{n}(\widehat{\log \mathrm{RV}}_{t+1|t} - \log \mathrm{RV}_{t+1})^2$ where $\widehat{\log \mathrm{RV}}_{t+1|t}$ is the prediction obtained by either the HAR model or the lasso and n is the total number of out-of-sample predictions. Tab. 4 shows two points: First, both the lasso and the HAR need a certain window length to attain reasonably low mean squared prediction errors (MSPEs), although the HAR model is markedly better for small training window sizes. Second, for longer training windows, the lasso and the HAR are almost equal in terms of MSPE.

To better understand these results we further report the evaluation over different out-of-sample periods: Pre-crisis, post-crisis, and full sample. The date for the beginning of the financial crisis was set to Sep 1, 2007. For the relevant training window lengths (i.e., 1,000 days and 2,000 days) we kept the maximal out-of-sample period which, unlike Tab. 4, results in evaluation windows of different lengths. The difference in MSPE is then tested using the Diebold-Mariano test (Diebold & Mariano 1995). These results are reported in Tab. 5.

Table 5: Diebold-Mariano (Diebold & Mariano 1995) tests of equal predictive ability

			AA	C	HAS	HDI	INTC	MSFT	NKE	PFE	XOM
1,000	Total	Mean Diff.	0.002	-0.001	0.001	-0.003	-0.001	-0.002	-0.006	**-0.007**	-0.001
	(n=1'383)	*p*-value	*0.38*	*0.86*	*0.85*	*0.34*	*0.49*	*0.35*	*0.18*	*0.00*	*0.48*
	PreCrisis	Mean Diff.	0.002	0.000	0.002	-0.006	0.001	-0.001	-0.012	**-0.008**	0.000
	(n=575)	*p*-value	*0.57*	*1.00*	*0.78*	*0.22*	*0.50*	*0.55*	*0.17*	*0.01*	*1.00*
	PostCrisis	Mean Diff.	0.002	-0.001	0.000	0.000	-0.003	-0.003	-0.001	**-0.006**	-0.003
	(n=808)	*p*-value	*0.49*	*0.84*	*0.96*	*0.94*	*0.26*	*0.44*	*0.69*	*0.05*	*0.39*
2,000	Total	Mean Diff.	0.002	0.000	-0.003	0.001	**0.002**	0.001	-0.002	0.003	0.002
	(n=383)	*p*-value	*0.38*	*0.92*	*0.44*	*0.72*	*0.09*	*0.62*	*0.30*	*0.18*	*0.33*
	PreCrisis	Mean Diff.	—	—	—	—	—	—	—	—	—
		p-value	—	—	—	—	—	—	—	—	—
	PostCrisis	Mean Diff.	0.002	0.000	-0.003	0.001	**0.002**	0.001	-0.002	0.003	0.002
	(n=383)	*p*-value	*0.38*	*0.91*	*0.44*	*0.72*	*0.09*	*0.62*	*0.29*	*0.18*	*0.33*

Difference in MSPE ($\text{MSPE}_{\text{HAR}} - \text{MSPE}_{\text{lasso}}$) are reported together with *p*-values from the Diebold-Mariano (Newey-West (Newey & West 1987) adjusted). The differences and *p*-values are reported for different training windows (1,000, 2,000) and before/after the financial crisis. Differences significant at 0.1 are typeset in boldface

Although there are a small number of rejections of the null we find no consistent pattern, neither in favor of the HAR nor in favor of the lasso. Investigating the predictions $\widehat{\log \text{RV}}^{(\text{lasso})}_{t+1|t}$ and $\widehat{\log \text{RV}}^{(\text{HAR})}_{t+1|t}$ in the sense of Mincer & Zarnowitz (1969) we find no evidence of either of the models (reported in Appendix E) being more often unbiased.

To be retained at this stage is that there is no clear evidence that either of the two models is genuinely better suited to forecast realized variance out-of-sample.

3.4 Risk Management Application

To test the predictions obtained from the lasso and the HAR model from a different angle, we include a risk management application. The value at risk of an asset to the level α is given as

$$\mathrm{VaR}_{\alpha}^{t} = -\inf\{x \in \mathbb{R} | \, \mathrm{P}(X_t \leq x) \geq 1 - \alpha\} \tag{9}$$

where X_t is the daily log-return of an asset.[8] Under the assumption, which also underlies the computation of realized variance, that an asset's return X_t is given as[9]

$$X_t = \mu_t + \sigma_t \cdot Z_t$$

we can readily compute (assuming a scale-location family with continuous distribution function) as

$$\mathrm{VaR}_{\alpha}(X) = \mu_t + \sigma_t q_{1-\alpha} \tag{10}$$

where $q_{1-\alpha}$ is the $1 - \alpha$ quantile of the standardized distribution Z_t, μ_t the conditional mean, and σ_t the conditional volatility of X.

As the distribution for Z_t we use the standard normal distribution as well as the empirical distribution after (quasi-)standardizing X_t with μ_t and estimates of σ_t as obtained by the RV_t estimates. Since we are aiming for a realistic benchmark we do not employ backtesting for the value at risk but conduct an out-of-sample analysis and predict

$$\mathrm{VaR}_{\alpha}^{t+1|t} = \mu_t + \sigma_{t+1|t} q_{1-\alpha} \tag{11}$$

where $\sigma_{t+1|t}$ is again obtained based on $\mathrm{RV}_{t+1|t}$ estimates by either the lasso or the HAR model.

To do so, we estimate both models on window lengths of $n = 200, 400, 1,000, 2,000$ observations to obtain a forecast $\widehat{\log \mathrm{RV}}_{t+1|t}$. To get an optimal forecast (in the sense of Proietti &

[8] We define the value-at-risk compliant to the risk management literature: Instead of working with the usual distribution, we premultipliy with -1 such that losses are positive resulting in the mnemonic that a greater VaR means greater risk.

[9] Strictly speaking the assumptions of computing realized variance also allow for jumps (depending on the estimator) to contribute to the return X_t. For reasons of simplicity, we exclude this component.

Lütkepohl (2011) and Appendix B) of the actual volatility we compute $\hat{\sigma}_{t+1|t}$ as

$$\hat{\sigma}_{t+1|t} = \sqrt{\exp(\widehat{\log \mathrm{RV}}_{t+1|t} - \frac{\tilde{\sigma}^2}{2})} \tag{12}$$

where $\tilde{\sigma}^2$ is the variance of $\log \mathrm{RV}_t$ and is computed as the empirical variance of $\{\widehat{\log \mathrm{RV}}_{t-n+1}, \ldots, \widehat{\log \mathrm{RV}}_t\}$ which produces $\hat{\sigma}^{(HAR)}_{t+1|t}$ and $\hat{\sigma}^{(lasso)}_{t+1|t}$. The same transformation is used to obtain the quantiles of the (quasi-)standardized residuals in (10).

The hit ratios are then defined as

$$\mathrm{HR}_{\alpha}^{M\,(D)} = \frac{\#\{x_{t+1} < -\mathrm{VaR}_{\alpha}^{t+1|t}\}}{n} \tag{13}$$

where 'M' can either be 'HAR' or 'lasso' depending on how $\sigma_{t+1|t}$ of (12) is computed (either by the HAR-model or our lasso approach), and 'D' is either 'Norm' or 'Emp' depending on how $q_{1-\alpha}$ in (11) is computed (quantiles of a $\mathcal{N}(0,1)$ distribution or quantiles of the standardized empirical distribution). In all cases we compute the conditional mean as $\mu_t = \frac{1}{n}\sum_{i=1}^{n} r_{t-n+i}$.

To contrast these estimates we also implement a naive estimator of the value at risk by simply taking the empirical α-quantile of the distribution of the log-returns, i.e.

$$\mathrm{HR}_{\alpha}^{Emp} = \frac{\#\{x_{t+1} < \hat{q}_{1-\alpha}\}}{n},$$

where $\hat{q}_{1-\alpha}$ is the empirical $1-\alpha$ quantile of $\{x_{t-n+1}, \ldots, x_t\}$.

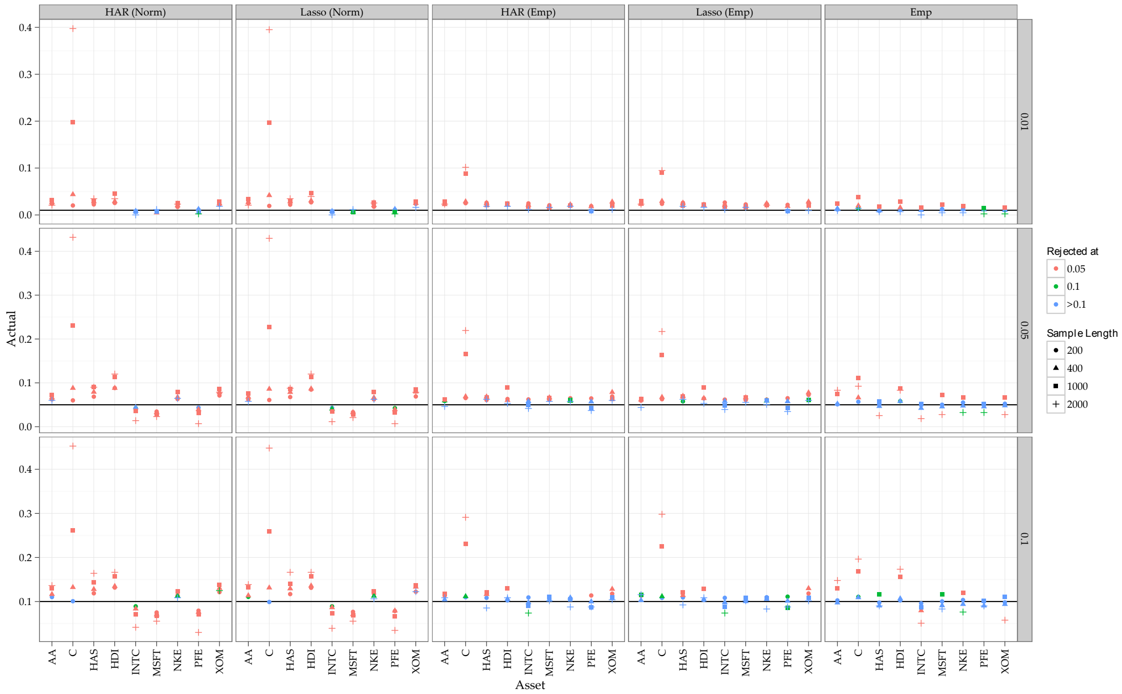

Figure 5: Actual hit ratios.The columns show the different estimators of HR_α, the rows show the levels of $\alpha = 99, 95, 90\%$. The horizontal lines are the theoretical levels $(1 - \alpha)$ of the VaR. The color indicates the p value of Kupiec's (1995) test against the theoretical level.

Fig. 5 clearly shows that there is again no systematic difference between the $\mathrm{HR}_\alpha^{HAR\ D}$ and $\mathrm{HR}_\alpha^{lasso\ D}$. Both are too aggressive (producing a VaR which is too low and thus is violated more often than theoretically specified) when the $\mathcal{N}(0, 1)$-distribution is used for the standardized innovations, and less so when the standardized empirical distribution is used. However, what becomes apparent from Fig. 5 is that the influence of the assumption on the distribution is much more crucial than the model used to forecast volatility. Compared to the simple model of estimating the VaR by simply taking the empirical quantiles the results are disappointing: There is no apparent outperformance of computing the VaR with volatility forecasts obtained by either the HAR or the lasso over simple (but effective) historical quantiles. This is all the more so, when looking at the rejections of $\mathcal{H}_0$ under Kupiec's (1995) test (assuming the correct level for the VaR). It is less often rejected for the 'Emp' than for any realized variance model.

The poor performance of all VaR forecasts for Citigroup, Inc. is related to the turbulent times the stock went through during the financial crisis resulting in pronounced non-normality of the $\log \mathrm{RV}_t$ as reported in Fig. 1 (b) as well as non-normality of the log-returns reported in Fig. 10.

4 Conclusions and Further Research

We conclude that the lasso does not recover the HAR model. We consider this as evidence against the presumption that HAR model is the true DGP since, first, we have theoretically founded reason to believe that the lasso should detect the HAR model, and, second, we provided empirical evidence on synthetic data that the lasso does recover the HAR model if the data stem from this DGP.

In addition, the lasso and the HAR model appear to be indistinguishable from an out-of-sample performance point of view: Neither the HAR nor the lasso excels in an out-of-sample prediction exercise. When we look at a more economically meaningful comparison using value at risk prediction, both models fare equally poorly with no noticeable differences in favor of either of the two.

The argument above and the selection of only near-lags (in the whole sample, and even more pronouncedly during the crisis) leads us to the hypothesis that in fact the realized variance dynamics are much better explained by shorter horizon models. Our results are in line withe empirical evidence shown in Chen, Härdle & Pigorsch (2010), eventually hinting at the possibility that the seemingly long-memory dynamics of the realized variance time series are in fact spurious. Arguments against this view are the lags which are selected and persist: This actually indicates that there might be some long range dependence which warrants further research.

We thus conclude that the HAR model may not be the true model. However, it captures – as does the lasso – a linear footprint of the possibly non-linear volatility dynamics that can be used for volatility forecasting. Given the equal out-of-sample performance of the two approaches we see potential for further research in this domain: Although adding additional predictors other than the lagged values of the realized volatilities themselves expels us from the thorough theoretical model selection framework established in this paper, we anticipate further insights with regard to e.g., volatility spillovers (including other assets, markets, etc. as predictors) or calendar effects (adding day-of-week dummies to the lasso regression).

References

Andersen, T., Dobrev, D. & Schaumburg, E. (2009), 'Jump-Robust Volatility Estimation using Nearest Neighbor Truncation', *NBER Working Paper No. 15533* .

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Ebens, H. (2001), 'The Distribution of Realized Stock Return Volatility', *Journal of Financial Economics* **61**(1), 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2001), 'The Distribution of Realized Exchange Rate Volatility', *Journal of the American Statistical Association* **96**(453), 42–55.

Areal, N. M. P. C. & Taylor, S. J. (2002), 'The Realized Volatility of FTSE-100 Futures Prices', *Journal of Futures Markets* **22**(7), 627–648.

Arlot, S. & Celisse, A. (2010), 'A Survey of Cross-validation Procedures for Model Selection', *Statistics Surveys* **4**, 40–79.

Barndorff-Nielsen, O. & Shephard, N. (2002), 'Econometric Analysis of Realized Volatility and its Use in Estimating Stochastic Volatility Models', *Journal of the Royal Statistical Society. Series B, Statistical Methodology* pp. 253–280.

Bergmeir, C. & Benítez, J. (2012), 'On the Use of Cross-validation for Time Series Predictor Evaluation', *Information Sciences* **191**, 192–213.

Bollerslev, T. (1986), 'Generalized Autoregressive Conditional Heteroskedasticity', *Journal of Econometrics* **31**, 307–327.

Box, G. & Cox, D. (1964), 'An Analysis of Transformations', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 211–252.

Brockwell, P. & Davis, R. (1986), *Time Series: Theory and Methods*, Springer-Verlag New York, Inc. New York, NY, USA.

Bühlmann, P. & Van De Geer, S. (2011), *Statistics for High-Dimensional Data: Methods, Theory and Applications*, Springer-Verlag New York Inc.

Chen, Y., Härdle, W. & Pigorsch, U. (2010), 'Localized Realized Volatility Modeling', *Journal of the American Statistical Association* **105**(492), 1376–1393.

Corsi, F. (2009), 'A Simple Approximate Long-Memory Model of Realized Volatility', *Journal of Financial Econometrics* **7**(2), 174–196.

Corsi, F., Audrino, F. & Renò, R. (2012), HAR Modelling for Realized Volatility Forecasting, *in* 'Handbook of Volatility Models and Their Applications', Wiley.

Craioveanu, M. & Hillebrand, E. (2010), Why it is OK to use the HAR-RV (1,5,21) Model, Technical report, Louisiana State University.

Diebold, F. X. & Mariano, R. S. (1995), 'Comparing Predictive Accuracy', *Journal of Business & Economic Statistics* **13**(3), 253–263.

Friedman, J., Hastie, T. & Tibshirani, R. (2010), 'Regularization Paths for Generalized Linear Models via Coordinate Descent', *Journal Of Statistical Software* **33**(1), 1–22.

Hansen, P. & Lunde, A. (2005), 'A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1, 1)?', *Journal of Applied Econometrics* **20**(7), 873–889.

Hart, J. D. & Wehrly, T. E. (1986), 'Kernel Regression Estimation Using Repeated Measurements Data', *Journal of the American Statistical Association* **81**(396), 1080–1088.

Hastie, T., Tibshirani, R. & Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*, 2nd edn, Springer.

Hintze, J. & Nelson, R. (1998), 'Violin Plots: A Box Plot-density Trace Synergism', *American Statistician* **52**(2), 181–184.

Hong-Zhi, A., Zhao-Guo, C. & Hannan, E. J. (1982), 'Autocorrelation, Autoregression and Autoregressive Approximation', *Annals of Statistics* **10**(3), 926–936.

Kock, A. (2012), On the Oracle Property of the Adaptive Lasso in Stationary and Nonstationary Autoregressions, Technical report, School of Economics and Management, University of Aarhus.

Kupiec, P. (1995), 'Techniques for Verifying the Accuracy of Risk Measurement Models', *Journal of Derivatives* **3**(2), 73–84.

Leeb, H. & Pötscher, B. (2005), 'Model Selection and Inference: Facts and Fiction', *Econometric Theory* **21**(1), 21–59.

Lewis, D. (1991), *Matrix Theory*, World Scientific Publishing Company Inc.

Martens, M., van Dijk, D. & de Pooter, M. (2009), 'Forecasting S&P 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements', *International Journal of Forecasting* **25**(2), 282–303.

McAleer, M. & Medeiros, M. C. (2008), 'Realized Volatility: A Review.', *Econometric Reviews* **27**(1-3), 10–45.

Mincer, J. & Zarnowitz, V. (1969), Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance, National Bureau of Economic Research, chapter The Evaluation of Economic Forecasts.

Nardi, Y. & Rinaldo, A. (2011), 'Autoregressive Process Modeling via the Lasso Procedure', *Journal of Multivariate Analysis* **102**(3), 528–549.

Newey, W. K. & West, K. D. (1987), 'A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix', *Econometrica* **55**(3), 703–708.

Pankratz, A. & Dudley, U. (1987), 'Forecasts of Power-transformed Series', *Journal of Forecasting* **6**(4), 239–239.

Park, H. & Sakaori, F. (2012), 'Lag Weighted Lasso for Time Series Model', *Computational Statistics* pp. 1–12. 10.1007/s00180-012-0313-5.

Proietti, T. & Lütkepohl, H. (2011), 'Does the Box-Cox transformation help in forecasting macroeconomic time series?', *OME Working Paper Series, University of Sydney* .

Thomakos, D. D. & Wang, T. (2003), 'Realized Volatility in the Futures Markets', *Journal of Empirical Finance* **10**(3), 321–353.

Tibshirani, R. (1996), 'Regression Shrinkage and Selection via the Lasso', *Journal of the Royal Statistical Society. Series B (Methodological)* **58**(1), 267–288.

Zhang, L., Mykland, P. & Aït-Sahalia, Y. (2005), 'A Tale of Two Time Scales', *Journal of the American Statistical Association* **100**(472), 1394–1411.

Zhao, P. & Yu, B. (2006), 'On Model Selection Consistency of Lasso', *Journal of Machine Learning Research* **7**(2), 2541–2563.

Zou, H. (2006), 'The Adaptive Lasso and Its Oracle Properties', *Journal of the American Statistical Association* **101**(476), 1418–1429.

A Proof of Theorem 2

This proof is structured as follows. We first show in Lemma 1 that the irrepresantable condition is satisfied for the HAR model. Based on this we invoke a theorem of Zhao & Yu (2006) which relaxes the assumptions on the innovation term for the lasso to be model consistent. Finally we show that the HAR model satisfies the assumptions of the aforementioned theorem and we can thus expect the lasso to be model selection consistent without the assumption Gaussianity for the error term.

Lemma 1. *Under the assumption that HAR model is true, condition (ii) of Theorem 1 is satisfied.*

Lemma 1 states that if the true DGP indeed obeys the law of motion as specified by the HAR model one can apply the results of Nardi & Rinaldo (2011) who establish that the lasso is a valid model selection device under two assumption, namely, that (i) $||\Gamma_{SS}^{-1}||_\infty \leq C$ and (ii) $||\Gamma_{S^cS}\Gamma_{SS}^{-1}||_\infty < 1$. Γ denotes the autocovariance matrix, S is the true active set of predictors, S^c is the true non-active set of predictors. When embedding the HAR model in this specification we have that S consists of the lagged values up to order 22 and S^c is any other lagged values beyond 22. Since (i) holds trivially as by (1) none of variables is a linear combination of another, we only collect the proof of (ii) in the Lemma below.

Proof. The proof is split into two parts. First we show that the infinity norm of $\Gamma_{S^cS}\Gamma_{SS}^{-1}$ can be seen as the sum of the absolute values of the regression coefficients of the usual HAR estimates, second, we show that it is sufficient to consider one specific non-active regressor.

Moreover, consider the following equivalent notations:

$$\mathrm{Cov}(S^c, S)\,\mathrm{Var}(S)^{-1} = \mathrm{Cov}(S^c, S)\,\mathrm{Cov}(S, S)^{-1} = \Gamma_{S^cS}\Gamma_{SS}^{-1}.$$

To rule out any possible confusion we re-state the definition of the infinity norm of a matrix. If $||\xi||_\infty$ for $\xi \in \mathbb{R}^n$ is defined as $||\xi||_\infty = \max_{1\leq i\leq n} |\xi_i|$, then the corresponding matrix norm is given as

$$||A||_\infty := \max_{||\xi||_\infty=1} ||A\xi||_\infty$$

where it can be shown (Lewis 1991, Proposition 3.4.1) for $A = [a_{ij}]_{1\leq i\leq n, 1\leq j\leq m}$ that

$$\|A\|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^{m} |a_{ij}|.$$

In what follows we consider a row-vector $\xi = [\xi_1, \ldots, \xi_n]$ as $1 \times n$ matrix such that $\|\xi\|_\infty = \|\xi'\|_1$.

Throughout the proof we assume without loss of generality the HAR model to contain no intercept. Moreover, for the sake of notational simplicity we assume the AR process to be labeled as

$$x_t = \sum_{i=1}^{22} \phi_i x_{t-i} + \varepsilon_t. \tag{14}$$

Assume that $|S^c| = 1$ with $S^c = \{x_{t-23}\}$[10] and that the true model is in fact the HAR model, i.e. $|S| = 22$ with $S = \{x_{t-1}, x_{t-2}, \ldots, x_{t-22}\}$. In other words, the active set consists of the first 22 lagged values and the first non-active predictor is x_{t-23}. We then find that

$$\mathrm{Cov}(x_{t-23}, [x_{t-1}, x_{t-2}, \ldots, x_{t-22}])\mathrm{Var}([x_{t-1}, x_{t-2}, \ldots, x_{t-22}])^{-1} = [\tilde{\phi}_1, \ldots, \tilde{\phi}_{22}], \tag{15}$$

where $[\tilde{\phi}_1, \ldots, \tilde{\phi}_{22}]$ is the usual representation of regression coefficients of x_{t-23} on $x_{t-1}, x_{t-2}, \ldots, x_{t-22}$ (note that the previously introduced superscript "HAR" is omitted to alleviate notation).

Since we are only interested in the sum of the absolute values of these regression coefficients, i.e. $\|[\tilde{\phi}_1, \ldots, \tilde{\phi}_{22}]\|_\infty$, we may as well reorder the regressors since

$$\|[\tilde{\phi}_1, \ldots, \tilde{\phi}_{22}]\|_\infty = \|[\tilde{\phi}_{\sigma(1)}, \ldots, \tilde{\phi}_{\sigma(22)}]\|_\infty \tag{16}$$

is true for any permutation σ. With $\sigma(i) = 22 - i + 1$ we find that

$$\|[\tilde{\phi}_{\sigma(1)}, \ldots, \tilde{\phi}_{\sigma(22)}]\|_\infty = \|\mathrm{Cov}(x_{t-23}, [x_{t-22}, x_{t-21}, \ldots, x_{t-1}])\mathrm{Var}([x_{t-22}, x_{t-21}, \ldots, x_{t-1}])^{-1}\|_\infty$$

A closer look at the second term (exploiting covariance stationarity and thus, the fact that the

[10]Observe that we slightly deviate from the notation used previously where $S \subset \mathbb{N}$; we use S and S^c to denote the corresponding lags variables rather than their indices.

autocovariance is an even function, (see for instance Brockwell & Davis (1986)) shows that

$$\begin{aligned}
\mathrm{Cov}(x_{t-23}, [x_{t-22}, x_{t-21}, \ldots, x_{t-1}] &= [\mathrm{Cov}(x_{t-23}, x_{t-(23-i)})]_{1\le i\le 22} \\
&= [\mathrm{Cov}(x_t, x_{t-i})]_{1\le i\le 22} \\
&= \mathrm{Cov}(x_t, [x_{t-1}, x_{t-2}, \ldots, x_{t-22}])
\end{aligned}$$

and

$$\mathrm{Var}([x_{t-22}, x_{t-21}, \ldots, x_{t-1}]) = \mathrm{Var}([x_{t-1}, x_{t-2}, \ldots, x_{t-22}])$$

such that

$$\begin{aligned}
[\tilde{\phi}_{\sigma(1)}, \tilde{\phi}_{\sigma(2)}, \ldots, \tilde{\phi}_{\sigma(22)}] &= \mathrm{Cov}(x_{t-23}, [x_{t-22}, x_{t-21}, \ldots, x_{t-1}])\,\mathrm{Var}([x_{t-22}, x_{t-21} \ldots, x_{t-1}) \\
&= \mathrm{Cov}(x_t, [x_{t-1}, x_{t-2}, \ldots, x_{t-22}])\,\mathrm{Var}([x_{t-1}, x_{t-2}, \ldots, x_{t-22}])^{-1} \\
&= [\phi_1, \phi_2, \ldots, \phi_{22}] \qquad (17)
\end{aligned}$$

Combining (16) and (17) shows that (15) is indeed simply the sum of the absolute values of the coefficients of (14), i.e., we conclude for $S^c = \{x_{t-23}\}$ that we have

$$\|\Gamma_{S^cS}\Gamma_{SS}^{-1}\|_\infty = \beta^{(d)} + \beta^{(w)} + \beta^{(m)}. \qquad (18)$$

When extending the set of non-active predictors to $S^C = \{x_{t-(22+i)}\}_{1\le i\le k}$ one can verify[11] that

$$\begin{aligned}
&\mathrm{Cov}([x_{t-(22+1)}, \ldots, x_{t-(22+k)]}, [x_{t-1}, x_{t-2}, \ldots, x_{t-22}])\mathrm{Var}([x_{t-1}, x_{t-2}, \ldots, x_{t-22}])^{-1} \\
&\qquad = \begin{bmatrix} \tilde{\phi}_1^{(1)} & \tilde{\phi}_2^{(1)} & \cdots & \tilde{\phi}_{22}^{(1)} \\ \vdots & \vdots & & \vdots \\ \tilde{\phi}_1^{(k)} & \tilde{\phi}_2^{(k)} & \cdots & \tilde{\phi}_{22}^{(k)} \end{bmatrix}. \qquad (19)
\end{aligned}$$

[11] This can either be seen by establishing the usual AR(p) moment conditions or recalling the fact that the OLS estimates of an AR(p) process are consistent.Note that the consistency of the AR(p) estimates only gives results a.s. by asymptotic equivalence. However, basing the argument on theoretical moments and the fact that for appropriate random matrices X and Y we have $[\mathrm{Cov}(Y, X)\,\mathrm{Var}(X)^{-1}]' = \mathrm{Var}(X)^{-1}\,\mathrm{Cov}(X, Y)$ yields (19) directly.

Hence,

$$\|\mathrm{Cov}(S^c, S)\mathrm{Var}(S)^{-1}\|_\infty = \max_{1\le j\le k} \sum_{i=1}^{22} |\tilde{\phi}_i^{(j)}|.$$

In a next step we show that $\sum_{i=1}^{22} |\phi_i^{(l)}| < \sum_{i=1}^{22} |\phi_i^{(k)}|$ for $l > k$ by induction. The conclusion then follows since it holds for $k = 1$, i.e. for $S^c = x_{t-23}$ which has already been proved in (18).

Given the argument which shows that reversing the order has no effect on the sum of the coefficients we present the argument in the usual AR(22) representation as given in (14) and thus drop the tilde, i.e.

$$x_{t+j} = \sum_{i=1}^{22} \phi_i^{(j)} x_{t+j-i} + \varepsilon_{t+j}.$$

Now, consider the induction basis for $j = 1 \to 2$:

$$\begin{aligned}
x_{t+1} &= \sum_{i=1}^{22} \phi_i^{(1)} x_{t+1-i} + \epsilon_{t+1} \\
&= \phi_1^{(1)} \left(\sum_{i=1}^{22} \phi_i^{(1)} x_{t-i} + \epsilon_t \right) + \sum_{i=2}^{22} \phi_i^{(1)} x_{t+1-i} + \epsilon_{t+1} \\
&= \sum_{i=1}^{21} \left(\phi_1^{(1)} \phi_i^{(1)} + \phi_{i+1}^{(1)} \right) x_{t-i} + \phi_1^{(1)} \phi_{22}^{(1)} x_{t-22} + \tilde{\epsilon}_{t+1} \\
&= \sum_{i=1}^{22} \phi_i^{(2)} x_{t-i} + \tilde{\epsilon}_{t+1},
\end{aligned}$$

where $\tilde{\epsilon}_{t+1} = \phi_1^{(1)} \epsilon_t + \epsilon_{t+1}$ and

$$\phi_i^{(2)} = \phi_1^{(1)} \phi_i^{(1)} + \phi_{i+1}^{(1)} \text{ for } i = 1, \ldots, 21 \text{ and } \phi_{22}^{(2)} = \phi_1^{(1)} \phi_{22}^{(1)}. \tag{20}$$

By the assumptions put forward in (1) we have that $\phi_i^{(2)} > 0 \ \forall i = 1, \ldots, 22$ and taking the difference of the sum of absolute values thus yields

$$\sum_{i=1}^{22} |\phi_i^{(2)}| - \sum_{i=1}^{22} |\phi_i^{(1)}| = \phi_1^{(1)} \left(\sum_{i=1}^{22} \phi_i^{(1)} - 1 \right) = \phi_1^{(1)} \left(\beta^{(d)} + \beta^{(w)} + \beta^{(m)} - 1 \right).$$

By (20) and (3) we have the induction basis $\phi_i^{(2)} > 0 \ \forall i = 1 \dots 22$ and also we find by the fact[12] $\beta^{(d)} + \beta^{(w)} + \beta^{(m)} < 1$ that $\sum_{i=1}^{22} \phi_i^{(j-1)} < \sum_{i=1}^{22} \phi_i^{(j)}$.

Reapplying the same argument for the induction step $j \to j+1$ yields

$$\begin{aligned}
x_{t+j} &= \sum_{i=1}^{22} \phi_i^{(j)} x_{t+1-i} + \varepsilon_{t+j} \\
&= \phi_1^{(j)} \left(\sum_{i=1}^{22} \phi_i^{(1)} x_{t-i} + \varepsilon_t \right) + \sum_{i=2}^{22} \phi_i^{(j)} x_{t+1-i} + \varepsilon_{t+j} \\
&= \sum_{i=1}^{21} \left(\phi_1^{(j)} \phi_i^{(1)} + \phi_{i+1}^{(j)} \right) x_{t-i} + \phi_{22}^{(1)} \phi_1^{(j)} x_{t-22} + \tilde{\varepsilon}_{t+j} \\
&= \sum_{i=1}^{22} \phi_i^{(j+1)} x_{t-i} + \tilde{\varepsilon}_{t+j}
\end{aligned}$$

where again $\tilde{\varepsilon}_{t+j} = \phi_1^{(1)} \varepsilon_t + \varepsilon_{t+j}$ and $\phi_i^{(j+1)} = \phi_1^{(j)} \phi_i^{(1)} + \phi_{i+1}^{(j)}$ for $i = 1, \dots, 21$ and $\phi_{22}^{(j+1)} = \phi_1^{(1)} \phi_{22}^{(j)}$.

Taking the difference between the sum of $\phi_i^{(j+1)}$ and the sum of $\phi_i^{(j)}$ yields

$$\sum_{i=1}^{22} \phi_i^{(j+1)} - \sum_{i=1}^{22} \phi_i^{(j)} = \left(\sum_{i=1}^{22} \phi_i^{(1)} - 1 \right) \phi_1^{(j)}.$$

By the induction basis we have $\phi_i^{(j)} > 0 \ \forall i = 1, \dots, 22$ such that $\phi_i^{(j+1)} > 0 \ \forall i = 1, \dots, 22$ and thus

$$\sum_{i=1}^{22} |\phi_i^{(j+1)}| - \sum_{i=1}^{22} |\phi_i^{(j)}| < 0$$

such that the claim

$$\sum_{i=1}^{22} |\phi_i^{(j+1)}| < \sum_{i=1}^{22} |\phi_i^{(j)}|$$

follows. Summarizing we conclude that for the HAR model it holds that $||\Gamma_{S^cS}\Gamma_{SS}||_\infty \leq 1 - \delta$ if $\beta^{(d)} + \beta^{(w)} + \beta^{(m)} \leq 1 - \delta$. □

Having proven the above we look at a theorem provided by Zhao & Yu (2006) which shows that the lasso is model selection consistent under some assumptions. Later we will prove that these assumptions hold if the HAR model is assumed to be true and we can thus safely relax

[12]This follows directly from the causality assumption: Since all roots lie outside the unit circle and the $P(z)$, the characteristic polynomial, is continous on $\mathbb{R}$ it follows that $P(1) > 0$ and thus that $\beta^{(d)} + \beta^{(w)} + \beta^{(m)} < 1$.

the assumption of normally distributed errors if we are willing to accept a fixed S and S^c (as opposed to Nardi & Rinaldo's (2011) results where $p = |S|$ is allowed to grow as the sample size increases.

Theorem A (Zhao & Yu (2006))**.** *Under the assumptions of S and S^c fixed and*

(A1) $|\Gamma_{S^cS}\Gamma_{SS}^{-1}\operatorname{sgn}(\operatorname{supp}\phi^0)| \overset{a.s.}{<} \mathbf{1}$ *where* $\mathbf{1}$ *is a vector of ones and the inequality is understood componentwise*

(A2) $\Gamma^n_{(S,S^c),(S,S^c)} \xrightarrow{a.s.} \Gamma_{(S,S^c),(S,S^c)}$ *where* $\Gamma_{(S,S^c)}$ *is the autocovariance matrix and* $\Gamma^n_{(S,S^c)}$ *its sample analogon*

(A3) $\frac{1}{n}\max_{0\le i\le n-p}\sum_{j=1}^{p} x^2_{t-i-j} \xrightarrow{a.s.} 0$

the lasso is model selection consistent in the sense of Definition 1 if the innovation term has finite second moment and λ_n *is chosen such that* $\lambda_n/n \to 0$ *and* $\lambda_n/n^{\frac{1+c}{2}} \to \infty$ *with* $0 \le c < 1$.

Proof of Theorem 2. We prove that the assumptions of Theorem A above are satisfied if one assumes the dynamics of the HAR model as put forward in (1) to hold as well as the existence of a finite fourth moment of the innovation term.

(A1) $|\Gamma_{S^cS}\Gamma_{SS}^{-1}\operatorname{sgn}(\operatorname{supp}\phi^0)| \overset{a.s.}{<} \mathbf{1}$ in (A1) of Theorem A holds since the argument in the proof of Lemma 1 can be made in terms of sample moments. Knowing that the least squares estimates converge a.s. to the true values (Brockwell & Davis 1986, Theorem 10.8.1) the conclusion follows since $|\Gamma_{S^cS}\Gamma_{SS}^{-1}\operatorname{sgn}(\operatorname{supp}\phi^0)| \overset{a.s.}{<} \mathbf{1}$ is weaker than $\|\Gamma_{S^cS}\Gamma_{SS}^{-1}\|_\infty \le 1-\delta$ as all components of $\operatorname{supp}\phi^0$ are greater than zero by (3).

(A2) Under the assumption of a finite fourth moment of the innovations we have by a result of Hong-Zhi, Zhao-Guo & Hannan (1982) the convergence almost surely. The positive definiteness follows from the fact that $\Gamma_{(S,S^c)}$ is positive semi-definite iff a variable is a linear combination of the others which is ruled out by the assumption of the HAR model as given in (3).[13]

(A3) Assuming that x_i is finite almost surely gives that $\frac{1}{n}\max_{0\le i\le n-p}\sum_{j=1}^{p} x^2_{t-i-j}$ is of class $\mathbf{o}_{a.s.}(n)$.

The condition on the innovation follows from Hölder's inequality since we have that $L^4 \subset L^2$ such that it suffices to require a finite fourth moment of the error term. □

[13] It is semi-definite since it is a covariance matrix.

Summarizing we have that the lasso should detect the HAR model if we assume a finite fourth moment.

B Log-Transformed Volatilities

Although it is common to use the log-transform to model realized variance for reasons of positiveness, lower skewness and lower kurtosis, the case of the HAR model even allows for additional arguments to justify the use of log-transformed realized volatilities. These are not solely related to the realized volatility series as such (as for instance in Martens, van Dijk & de Pooter (2009, Table 1)) but also to how realized volatility is modeled. Extending the approach of Box & Cox (1964) where only the dependent variable is transformed we employ the Box-Cox transform

$$f_\lambda(x) = x^{(\lambda)} = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(x) & \text{otherwise.} \end{cases}$$

to series of realized volatility. Consequently, the Box-Cox transform not only affects the dependent variable but also predictor variables in the HAR model. As in the original work of Box & Cox we then compute the (quasi-)likelihood for each λ. Since the (quasi-)likelihood is equivalent to the R^2 we report the R^2 for different values of λ in Fig. 6.

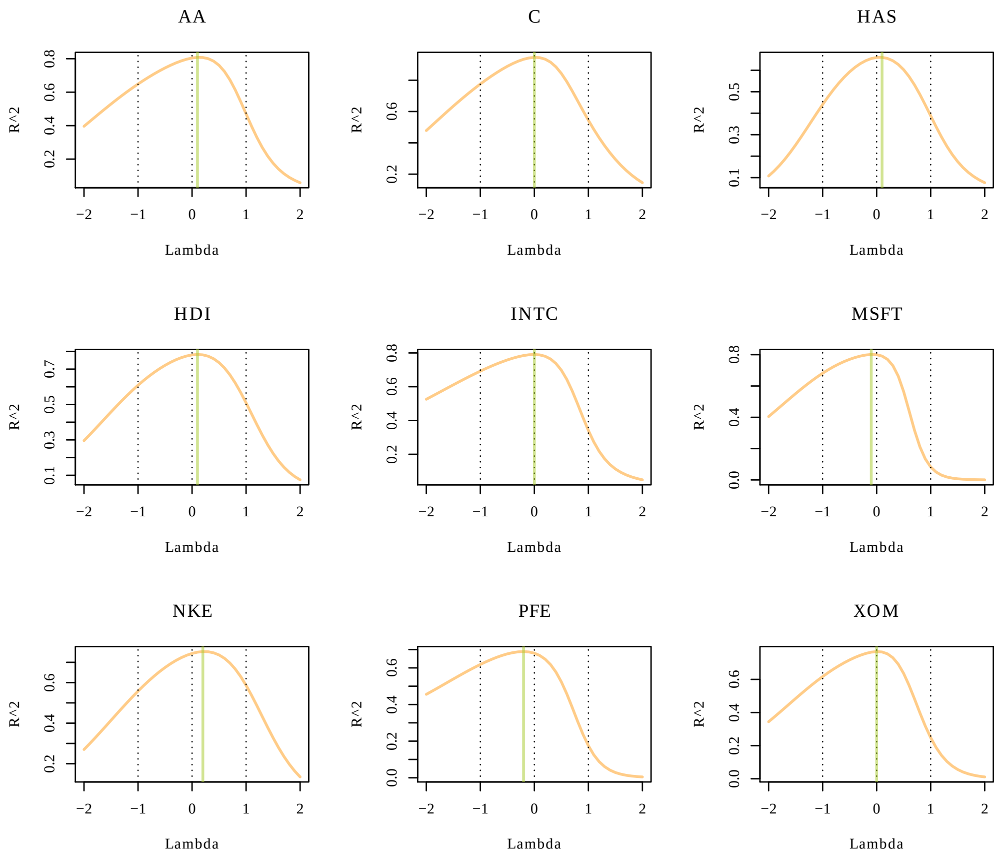

Figure 6: R^2 for different values of λ for the HAR model estimated on $\mathrm{RV}_t^{(\lambda)}$ on the whole sample as described in Section 3.1. The green line indicates the maximal R^2 and the dotted lines indicate common transformations for realized volatilities ($\sqrt{\mathrm{RV}_t}$ with $\lambda = -1$, $\log \mathrm{RV}_t$ with $\lambda = 0$, and RV_t with $\lambda = 1$)

Clearly, following again Box & Cox and choosing a rational λ it follows that $\lambda = 0$ is a sensitive choice and thus justifies the use of log-transformed volatilities. A further argument for using $\lambda = 0$ may be found in the fact that for the case of $\lambda = 0$ we can construct unbiased estimates (under the assumption of normality of the log-transformed realized volatilities) explicitly without resorting to the median (Pankratz & Dudley 1987, Proietti & Lütkepohl 2011).

C Robustness

This section shows the key results in graphical form as presented in the main paper if the realized volatility is estimated by Andersen et al.'s (2009) MedRV estimator instead of Zhang et al.'s (2005) two-time-scale estimator. MedRV is not only computationally attractive but also robust to zero returns and outliers induced by jumps. Figures 7 to 9 and Tab. 6 are found below and are otherwise identical to the corresponding figures in the main text. There are marginal differences, but, the conclusions made in the main text remain valid such that we abstain from further discussion of these results.

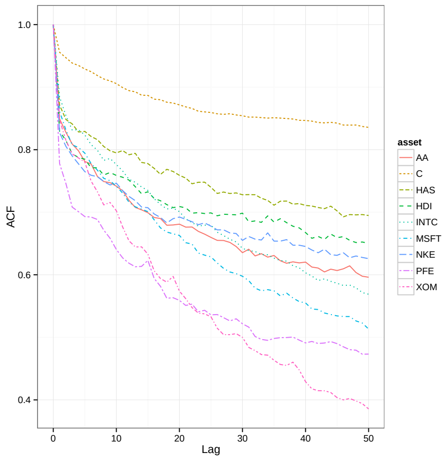

Figure 7 (a)

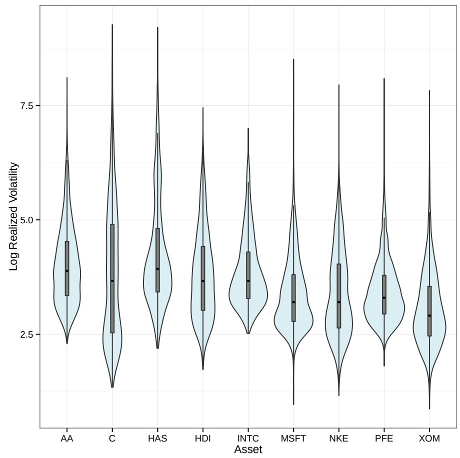

Figure 7 (b)

Figure 7: Panel (a) shows the autocorrelation function for the 9 $\log \mathrm{RV}_t$ series. Panel (b) shows a violin plot (Hintze & Nelson 1998) of the unconditional $\log \mathrm{RV}_t$. Both use the MedRV estimator.

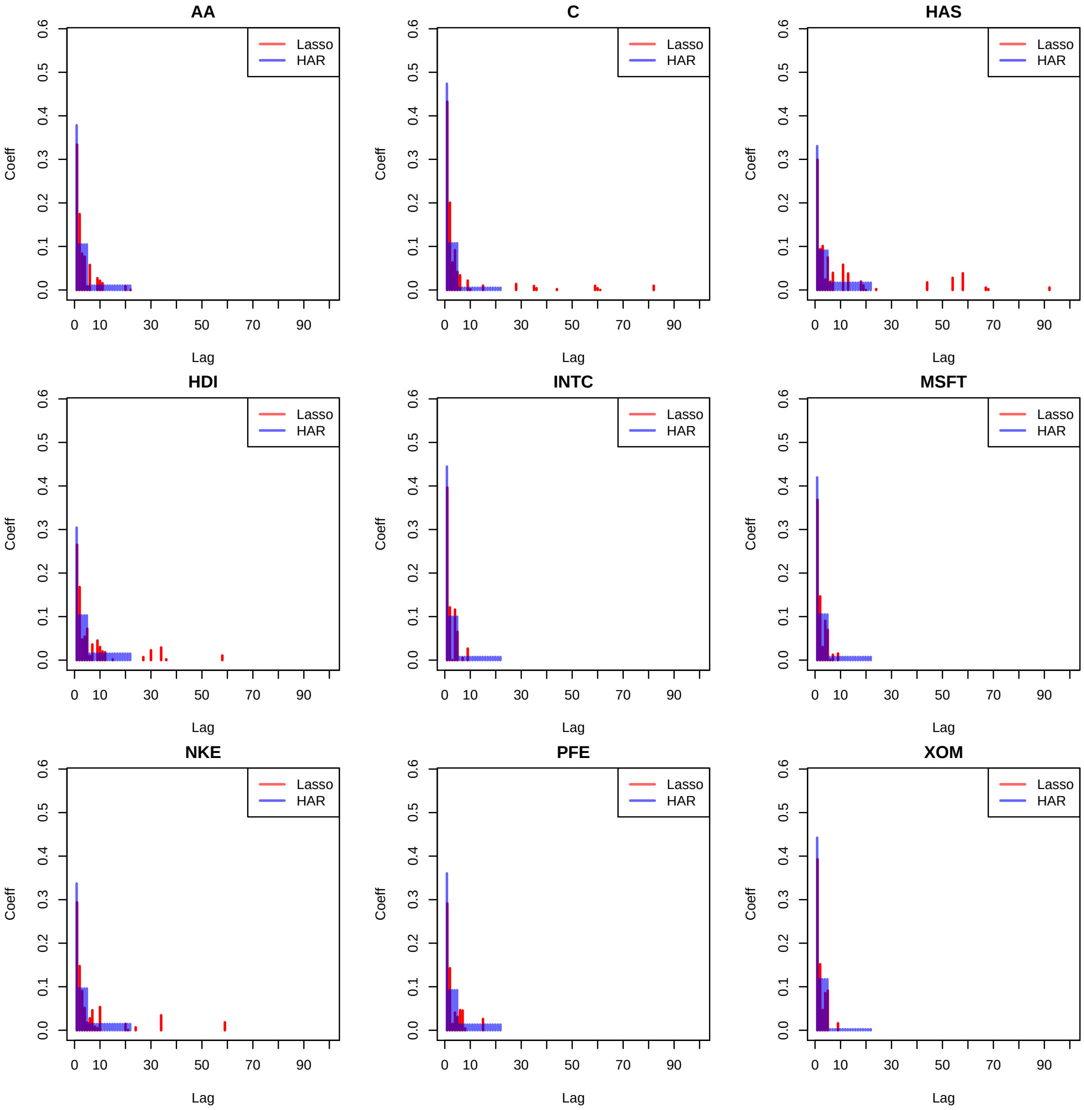

Figure 8: HAR versus lasso coefficients with all predictors using MedRV estimator

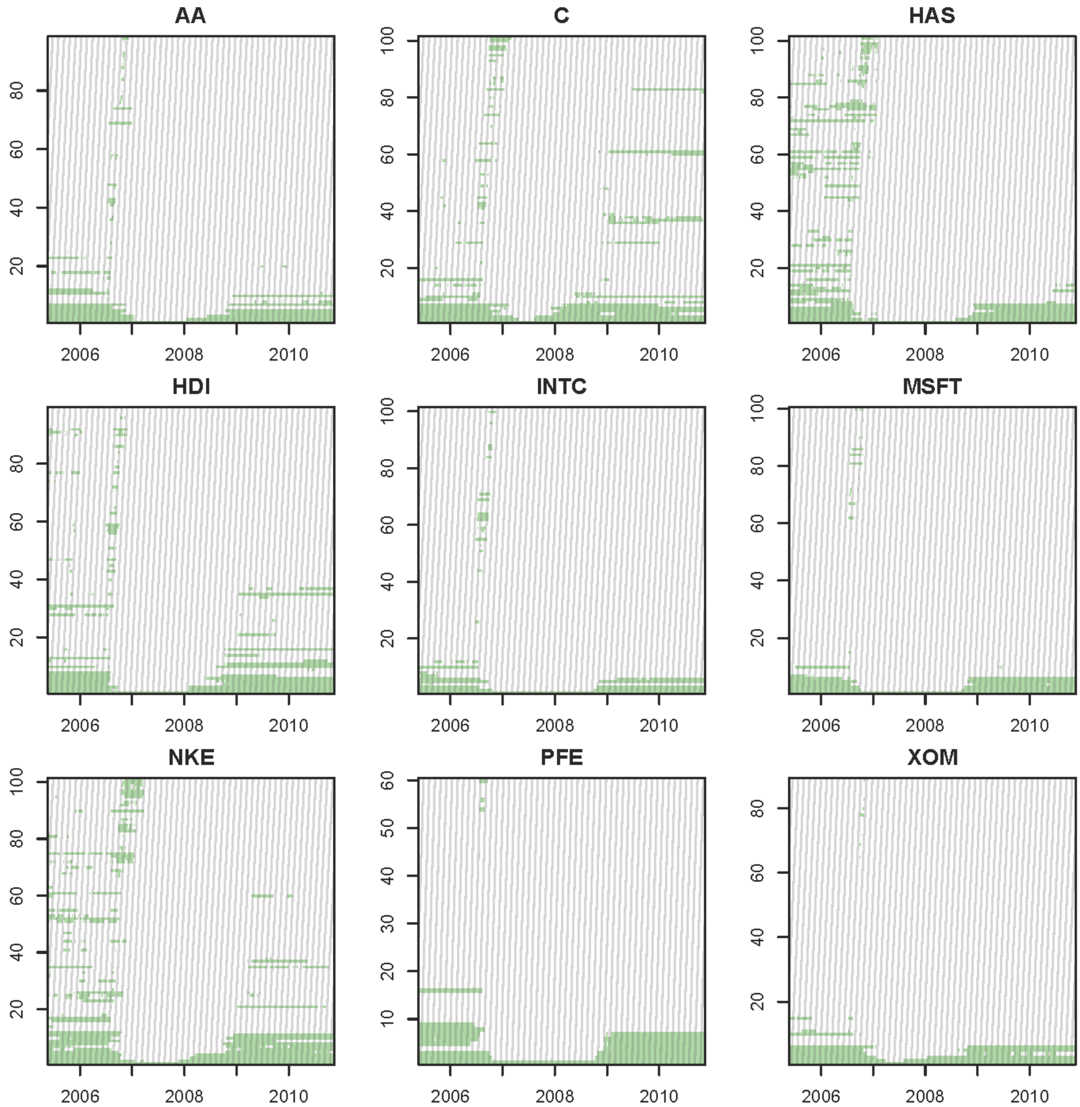

Figure 9: Stability of Lasso selected Regressors for all assets using MedRV estimator

Table 6: Diebold-Mariano (Diebold & Mariano 1995) tests of equal predictive ability

			AA	C	HAS	HDI	INTC	MSFT	NKE	PFE	XOM
1,000	Total	Mean Diff.	0.000	0.000	**-0.005**	-0.001	-0.001	**-0.002**	-0.004	**-0.003**	0.000
	(n=1'383)	p-value	*1.00*	*0.86*	*0.03*	*0.66*	*0.52*	*0.05*	*0.25*	*0.01*	*0.91*
	PreCrisis	Mean Diff.	-0.001	-0.003	-0.006	-0.002	0.000	-0.001	-0.005	-0.003	0.000
	(n=575)	p-value	*0.87*	*0.25*	*0.13*	*0.71*	*0.89*	*0.32*	*0.51*	*0.11*	*0.90*
	PostCrisis	Mean Diff.	0.000	0.003	-0.004	-0.001	-0.001	**-0.003**	-0.004	**-0.003**	0.000
	(n=808)	p-value	*0.88*	*0.48*	*0.13*	*0.81*	*0.40*	*0.08*	*0.29*	*0.06*	*0.96*
2,000	Total	Mean Diff.	0.001	-0.001	-0.003	0.002	0.000	0.001	0.000	0.001	0.003
	(n=383)	p-value	*0.70*	*0.21*	*0.39*	*0.57*	*0.82*	*0.61*	*0.95*	*0.56*	*0.31*
	PreCrisis	Mean Diff.	—	—	—	—	—	—	—	—	—
	(n=—)	p-value	—	—	—	—	—	—	—	—	—
	PostCrisis	Mean Diff.	0.001	-0.001	-0.003	0.002	0.000	0.001	0.000	0.001	0.003
	(n=383)	p-value	*0.70*	*0.21*	*0.39*	*0.57*	*0.82*	*0.61*	*0.95*	*0.56*	*0.31*

Difference in MSPE ($MSPE_{HAR} - MSPE_{lasso}$) are reported together with p-values from the Diebold-Mariano (Newey-West (Newey & West 1987) adjusted). The differences and p-values are reported for different training windows (1,000, 2,000) and before/after the financial crisis using the MedRV estimator.

D Risk Management Application

This section contains the actual violations of the value at risk visualized in Fig. 5 collected in Tab. 8. Moreover, we have added summary statistics for the distribution of returns in Fig. 10.

Table 7: Hit Ratios

		α	AA Act	AA Exp	AA p Val	C Act	C Exp	C p Val	HAS Act	HAS Exp	HAS p Val	HDI Act	HDI Exp	HDI p Val	INTC Act	INTC Exp	INTC p Val	MSFT Act	MSFT Exp	MSFT p Val	NKE Act	NKE Exp	NKE p Val	PFE Act	PFE Exp	PFE p Val	XOM Act	XOM Exp	XOM p Val
200	Emp	0.01	26	22	0.45	31	22	0.08	19	22	0.47	26	22	0.45	25	22	0.58	28	22	0.25	27	22	0.34	27	22	0.34	22	22	0.94
	HAR (Emp)	0.01	55	22	0.00	56	22	0.00	58	22	0.00	55	22	0.00	55	22	0.00	46	22	0.00	48	22	0.00	42	22	0.00	56	22	0.00
	HAR (Norm)	0.01	53	22	0.00	45	22	0.00	50	22	0.00	57	22	0.00	16	22	0.16	16	22	0.16	38	22	0.00	25	22	0.58	49	22	0.00
	lasso (Emp)	0.01	53	22	0.00	53	22	0.00	58	22	0.00	51	22	0.00	59	22	0.00	50	22	0.00	52	22	0.00	48	22	0.00	59	22	0.00
	lasso (Norm)	0.01	57	22	0.00	43	22	0.00	49	22	0.00	60	22	0.00	15	22	0.10	18	22	0.34	39	22	0.00	25	22	0.58	53	22	0.00
	Emp	0.05	114	112	0.82	127	112	0.14	114	112	0.82	130	112	0.08	110	112	0.87	112	112	0.97	123	112	0.28	114	112	0.82	117	112	0.61
	HAR (Emp)	0.05	131	112	0.07	146	112	0.00	136	112	0.02	141	112	0.01	139	112	0.01	137	112	0.02	145	112	0.00	144	112	0.00	154	112	0.00
	HAR (Norm)	0.05	139	112	0.01	134	112	0.04	153	112	0.00	197	112	0.00	95	112	0.10	77	112	0.00	141	112	0.01	97	112	0.15	159	112	0.00
	lasso (Emp)	0.05	133	112	0.04	140	112	0.01	129	112	0.10	144	112	0.00	138	112	0.01	136	112	0.02	138	112	0.01	145	112	0.00	162	112	0.00
	lasso (Norm)	0.05	138	112	0.01	136	112	0.02	151	112	0.00	189	112	0.00	96	112	0.12	74	112	0.00	139	112	0.01	95	112	0.10	154	112	0.00
	Emp	0.1	229	223	0.69	247	223	0.10	218	223	0.71	229	223	0.69	214	223	0.51	224	223	0.96	231	223	0.59	224	223	0.96	219	223	0.76
	HAR (Emp)	0.1	253	223	0.04	245	223	0.13	242	223	0.19	229	223	0.69	244	223	0.15	236	223	0.37	239	223	0.27	254	223	0.03	263	223	0.01
	HAR (Norm)	0.1	246	223	0.11	225	223	0.90	265	223	0.00	294	223	0.00	199	223	0.08	168	223	0.00	254	223	0.03	177	223	0.00	271	223	0.00
	lasso (Emp)	0.1	253	223	0.04	242	223	0.19	244	223	0.15	231	223	0.59	242	223	0.19	235	223	0.41	245	223	0.13	248	223	0.09	264	223	0.01
	lasso (Norm)	0.1	247	223	0.10	221	223	0.87	261	223	0.01	293	223	0.00	199	223	0.08	171	223	0.00	262	223	0.01	175	223	0.00	273	223	0.00
400	Emp	0.01	28	20	0.11	40	20	0.00	20	20	0.94	32	20	0.02	21	20	0.88	23	20	0.56	24	20	0.43	23	20	0.56	21	20	0.88
	HAR (Emp)	0.01	46	20	0.00	58	20	0.00	51	20	0.00	43	20	0.00	43	20	0.00	36	20	0.00	45	20	0.00	37	20	0.00	57	20	0.00
	HAR (Norm)	0.01	55	20	0.00	89	20	0.00	58	20	0.00	59	20	0.00	17	20	0.44	10	20	0.01	41	20	0.00	25	20	0.31	49	20	0.00
	lasso (Emp)	0.01	47	20	0.00	60	20	0.00	52	20	0.00	42	20	0.00	43	20	0.00	39	20	0.00	50	20	0.00	38	20	0.00	57	20	0.00
	lasso (Norm)	0.01	55	20	0.00	85	20	0.00	58	20	0.00	62	20	0.00	16	20	0.32	13	20	0.08	42	20	0.00	25	20	0.31	51	20	0.00
	Emp	0.05	106	102	0.66	135	102	0.00	95	102	0.49	118	102	0.10	86	102	0.10	93	102	0.37	97	102	0.63	93	102	0.37	98	102	0.71
	HAR (Emp)	0.05	121	102	0.06	140	102	0.00	139	102	0.00	123	102	0.04	118	102	0.10	135	102	0.00	119	102	0.09	116	102	0.15	160	102	0.00
	HAR (Norm)	0.05	135	102	0.00	179	102	0.00	161	102	0.00	179	102	0.00	89	102	0.19	66	102	0.00	137	102	0.00	80	102	0.02	159	102	0.00
	lasso (Emp)	0.05	125	102	0.02	137	102	0.00	142	102	0.00	132	102	0.00	118	102	0.10	135	102	0.00	120	102	0.07	117	102	0.13	158	102	0.00
	lasso (Norm)	0.05	138	102	0.00	175	102	0.00	160	102	0.00	176	102	0.00	83	102	0.05	66	102	0.00	135	102	0.00	80	102	0.02	162	102	0.00
	Emp	0.1	198	203	0.69	222	203	0.17	184	203	0.15	217	203	0.32	162	203	0.00	185	203	0.17	191	203	0.36	184	203	0.15	191	203	0.36
	HAR (Emp)	0.1	212	203	0.52	227	203	0.08	237	203	0.01	215	203	0.39	199	203	0.75	211	203	0.57	221	203	0.20	203	203	0.98	261	203	0.00
	HAR (Norm)	0.1	237	203	0.01	269	203	0.00	260	203	0.00	274	203	0.00	170	203	0.01	143	203	0.00	229	203	0.06	155	203	0.00	265	203	0.00
	lasso (Emp)	0.1	209	203	0.67	227	203	0.08	232	203	0.04	212	203	0.52	199	203	0.75	204	203	0.96	214	203	0.43	206	203	0.84	264	203	0.00
	lasso (Norm)	0.1	231	203	0.04	267	203	0.00	263	203	0.00	276	203	0.00	176	203	0.04	144	203	0.00	229	203	0.06	161	203	0.00	271	203	0.00
1,000	Emp	0.01	35	14	0.00	54	14	0.00	26	14	0.01	40	14	0.00	23	14	0.03	32	14	0.00	27	14	0.00	21	14	0.10	23	14	0.03
	HAR (Emp)	0.01	41	14	0.00	127	14	0.00	30	14	0.00	34	14	0.00	25	14	0.01	23	14	0.03	29	14	0.00	12	14	0.52	28	14	0.00
	HAR (Norm)	0.01	45	14	0.00	284	14	0.00	44	14	0.00	65	14	0.00	10	14	0.22	10	14	0.22	36	14	0.00	9	14	0.13	40	14	0.00
	lasso (Emp)	0.01	43	14	0.00	129	14	0.00	29	14	0.00	31	14	0.00	25	14	0.01	24	14	0.02	30	14	0.00	12	14	0.52	29	14	0.00
	lasso (Norm)	0.01	48	14	0.00	282	14	0.00	43	14	0.00	67	14	0.00	10	14	0.22	8	14	0.07	38	14	0.00	8	14	0.07	40	14	0.00
	Emp	0.05	107	72	0.00	159	72	0.00	82	72	0.22	125	72	0.00	74	72	0.78	103	72	0.00	96	72	0.00	75	72	0.69	96	72	0.00
	HAR (Emp)	0.05	89	72	0.04	238	72	0.00	95	72	0.01	128	72	0.00	71	72	0.94	94	72	0.01	87	72	0.07	62	72	0.23	90	72	0.03
	HAR (Norm)	0.05	104	72	0.00	330	72	0.00	130	72	0.00	163	72	0.00	52	72	0.01	40	72	0.00	113	72	0.00	45	72	0.00	123	72	0.00
	lasso (Emp)	0.05	91	72	0.02	235	72	0.00	99	72	0.00	129	72	0.00	70	72	0.84	96	72	0.00	85	72	0.12	62	72	0.23	87	72	0.07
	lasso (Norm)	0.05	108	72	0.00	326	72	0.00	123	72	0.00	163	72	0.00	50	72	0.01	40	72	0.00	114	72	0.00	46	72	0.00	121	72	0.00
	Emp	0.1	187	143	0.00	242	143	0.00	166	143	0.05	223	143	0.00	125	143	0.10	166	143	0.05	171	143	0.02	146	143	0.81	158	143	0.20
	HAR (Emp)	0.1	169	143	0.03	330	143	0.00	173	143	0.01	186	143	0.00	129	143	0.20	159	143	0.17	150	143	0.56	125	143	0.10	157	143	0.23
	HAR (Norm)	0.1	187	143	0.00	375	143	0.00	205	143	0.00	226	143	0.00	102	143	0.00	97	143	0.00	177	143	0.00	101	143	0.00	197	143	0.00
	lasso (Emp)	0.1	165	143	0.06	323	143	0.00	173	143	0.01	184	143	0.00	126	143	0.12	156	143	0.27	153	143	0.40	123	143	0.07	155	143	0.31
	lasso (Norm)	0.1	189	143	0.00	372	143	0.00	201	143	0.00	226	143	0.00	105	143	0.00	98	143	0.00	176	143	0.01	96	143	0.00	196	143	0.00
2,000	Emp	0.01	4	4	0.87	6	4	0.45	3	4	0.50	3	4	0.50	0	4	NA	2	4	0.21	2	4	0.21	1	4	0.05	1	4	0.05
	HAR (Emp)	0.01	10	4	0.02	44	4	0.00	8	4	0.11	8	4	0.11	5	4	0.75	7	4	0.24	8	4	0.11	4	4	0.87	5	4	0.75
	HAR (Norm)	0.01	9	4	0.05	172	4	0.00	15	4	0.00	15	4	0.00	0	4	NA	5	4	0.75	10	4	0.02	1	4	0.05	8	4	0.11
	lasso (Emp)	0.01	10	4	0.02	41	4	0.00	8	4	0.11	7	4	0.24	5	4	0.75	7	4	0.24	9	4	0.05	4	4	0.87	4	4	0.87
	lasso (Norm)	0.01	9	4	0.05	171	4	0.00	15	4	0.00	17	4	0.00	0	4	NA	5	4	0.75	11	4	0.01	1	4	0.05	7	4	0.24
	Emp	0.05	36	22	0.00	40	22	0.00	11	22	0.01	36	22	0.00	8	22	0.00	12	22	0.02	14	22	0.07	14	22	0.07	12	22	0.02
	HAR (Emp)	0.05	20	22	0.71	95	22	0.00	27	22	0.26	23	22	0.77	18	22	0.41	25	22	0.47	25	22	0.47	16	22	0.19	26	22	0.35
	HAR (Norm)	0.05	26	22	0.35	187	22	0.00	39	22	0.00	52	22	0.00	6	22	0.00	10	22	0.00	28	22	0.18	3	22	0.00	34	22	0.01
	lasso (Emp)	0.05	19	22	0.55	94	22	0.00	27	22	0.26	23	22	0.77	17	22	0.29	24	22	0.61	22	22	0.94	15	22	0.12	26	22	0.35
	lasso (Norm)	0.05	25	22	0.47	186	22	0.00	38	22	0.00	52	22	0.00	5	22	0.00	9	22	0.00	27	22	0.26	3	22	0.00	34	22	0.01
	Emp	0.1	64	43	0.00	85	43	0.00	39	43	0.48	75	43	0.00	22	43	0.00	36	43	0.23	33	43	0.09	39	43	0.48	25	43	0.00
	HAR (Emp)	0.1	47	43	0.56	126	43	0.00	37	43	0.30	47	43	0.56	32	43	0.06	44	43	0.91	38	43	0.39	38	43	0.39	45	43	0.79
	HAR (Norm)	0.1	59	43	0.02	196	43	0.00	71	43	0.00	72	43	0.00	18	43	0.00	24	43	0.00	47	43	0.56	13	43	0.00	54	43	0.10
	lasso (Emp)	0.1	50	43	0.29	129	43	0.00	40	43	0.59	47	43	0.56	32	43	0.06	44	43	0.91	36	43	0.23	38	43	0.39	44	43	0.91
	lasso (Norm)	0.1	60	43	0.01	194	43	0.00	72	43	0.00	72	43	0.00	17	43	0.00	24	43	0.00	46	43	0.67	15	43	0.00	53	43	0.13

Table 8: VaR violations (hit ratios)

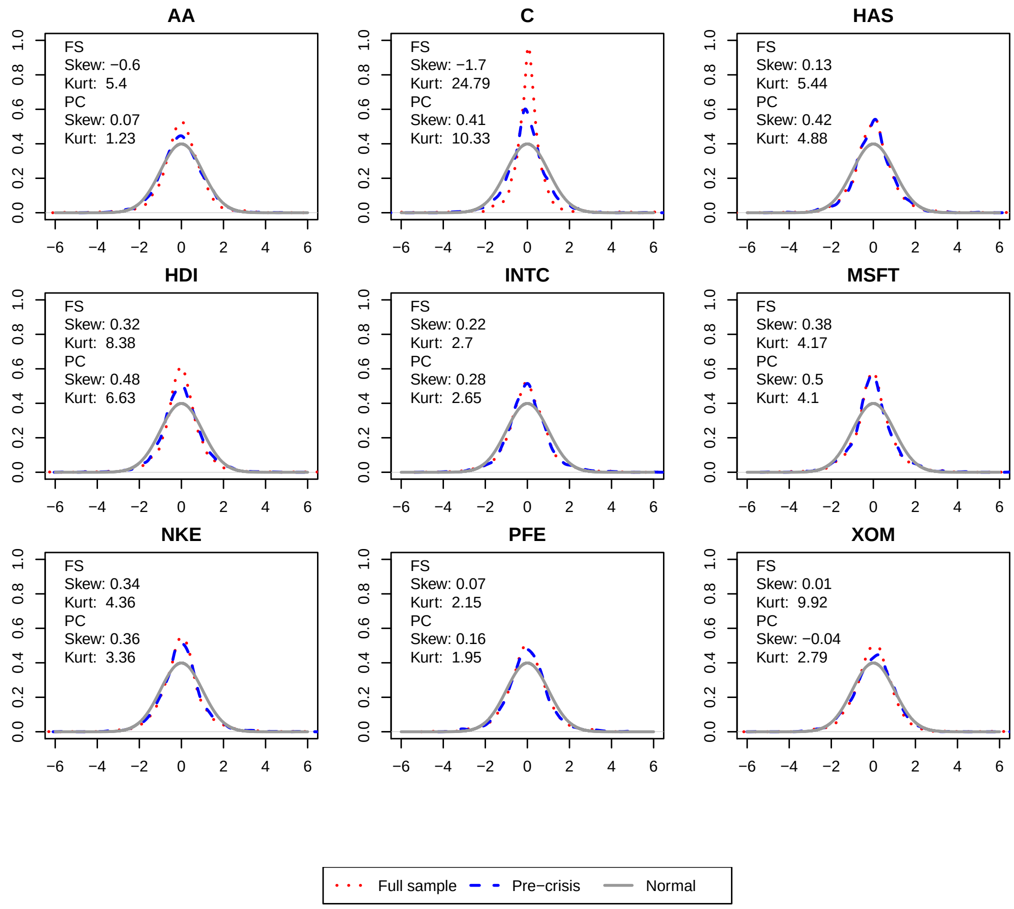

Figure 10: Kernel density estimates of standardized log-returns for pre-crisis (PC) and full sample (FS) against normal distribution.

E Mincer Zarnowitz Regressions

In this paragraph we present the Mincer-Zarnowitz (Mincer & Zarnowitz 1969) regressions for the lasso as well as the HAR model for the different training window lengths as well as split into pre-crisis (PrC), post-crisis (PoC), and full-sample (FS). Instead of reporting tables we include three figures: Fig. 11 contains the estimated intercept with 95% confidence intervals, Fig. 12 contains the estimated slope parameter with 95% confidence intervals, and Fig. 13 contains the

p-value of the joint hypothesis that the intercept equals 0 and the slope equals 1. Horizontal lines show the 5% and 10% level. In total the lasso is rejected 38 times (48 times) at the 5% level (10% level) whereas the HAR is rejected 50 times in both cases (out of 99 tests for each model). We account for dependence of the error term by using HAC consistent standard errors (Newey & West 1987).

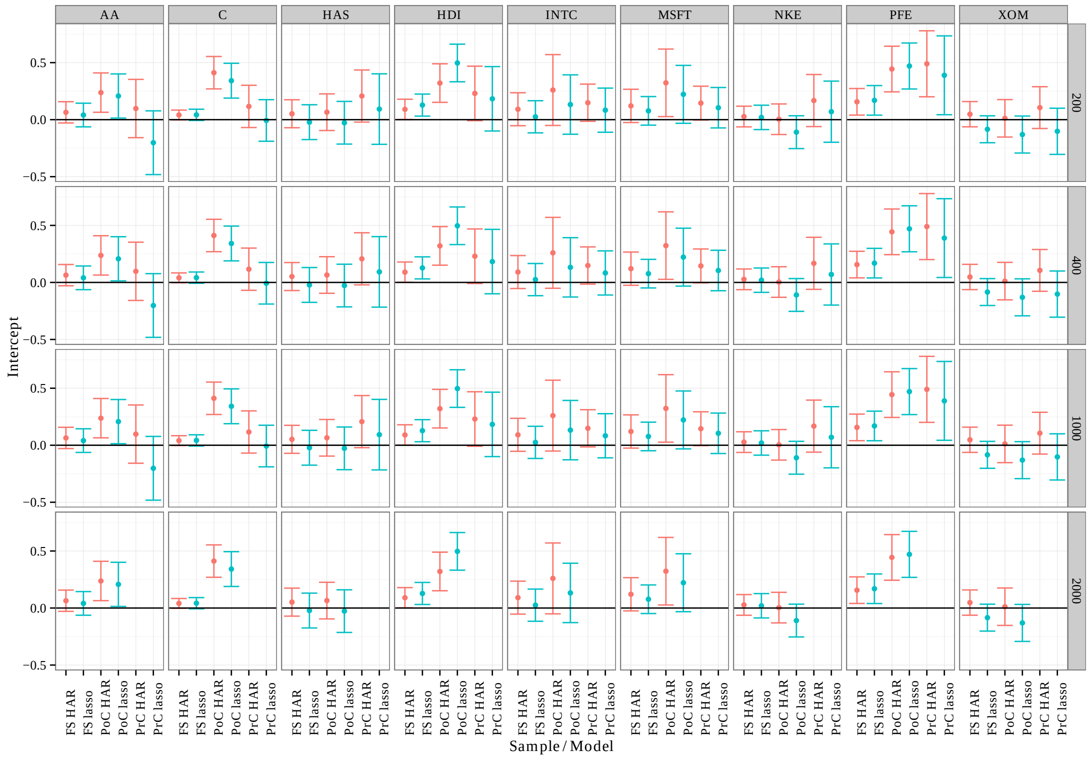

Figure 11: Estimate of $\hat{\alpha}$ in $\log \mathrm{RV}_t = \alpha + \beta \cdot \widehat{\log \mathrm{RV}}_t + \epsilon_t$

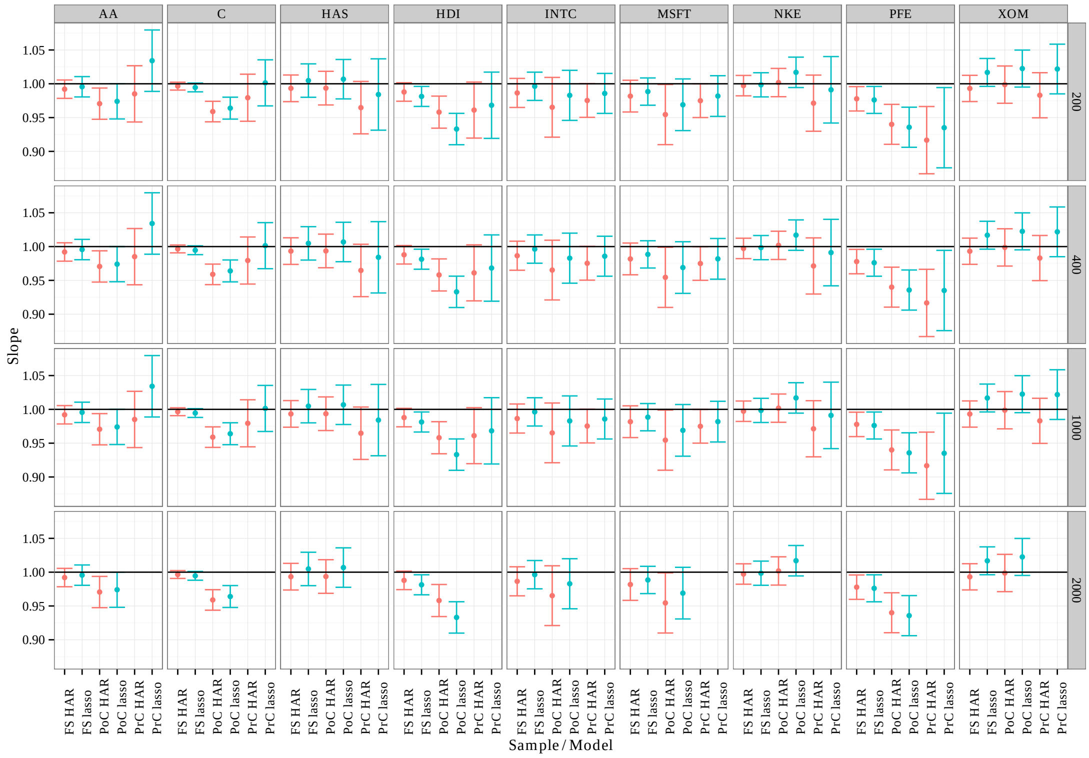

Figure 12: Estimate of $\hat{\beta}$ in $\log \mathrm{RV}_t = \alpha + \beta \cdot \widehat{\log \mathrm{RV}}_t + \epsilon_t$

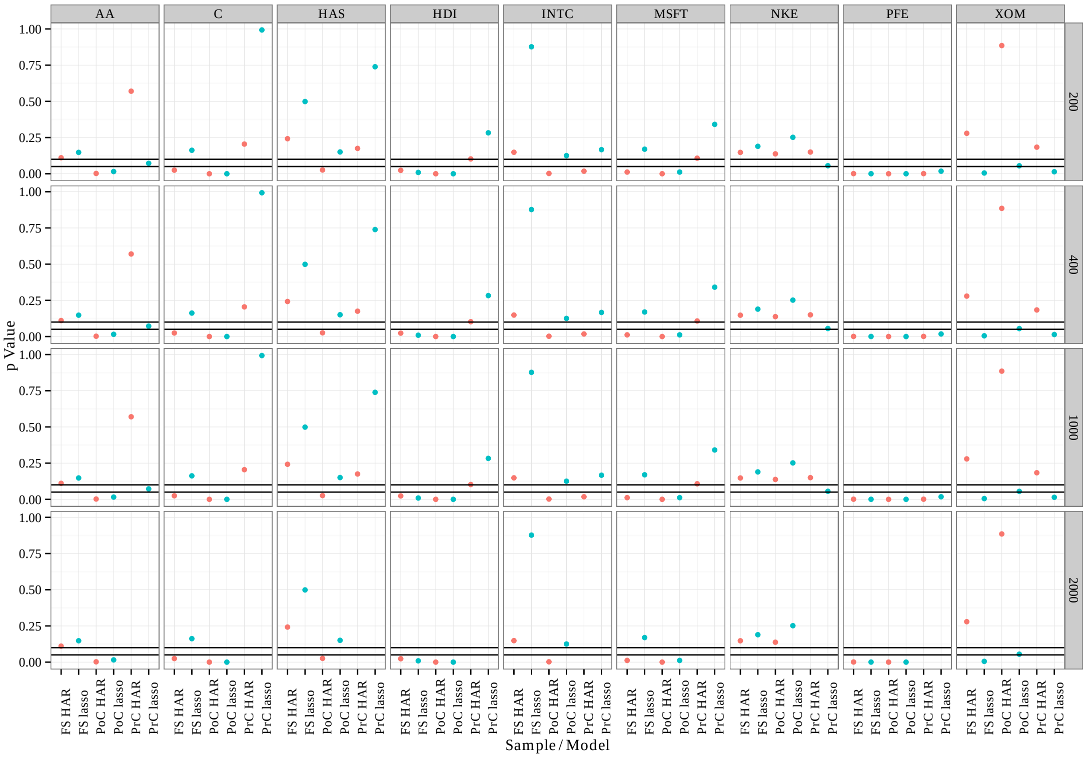

Figure 13: p-value $\mathcal{H}_0 : \alpha = 0 \wedge \beta = 1$ of $\log \mathrm{RV}_t = \alpha + \beta \cdot \widehat{\log \mathrm{RV}}_t + \epsilon_t$