



AARHUS UNIVERSITY



# Coversheet

---

## **This is the accepted manuscript (post-print version) of the article.**

Contentwise, the accepted manuscript version is identical to the final published version, but there may be differences in typography and layout.

## **How to cite this publication**

Please cite the final published version:

Callot, L., Kock, A. B., & Medeiros, M. (2017). Modeling and Forecasting Large Realized Covariance Matrices and Portfolio Choice. *Journal of Applied Econometrics*, 32(1), 140-158. <https://doi.org/10.1002/jae.2512>

## Publication metadata

<b>Title:</b>	Modeling and Forecasting Large Realized Covariance Matrices and Portfolio Choice
<b>Author(s):</b>	Callot, L., Kock, A. B., & Medeiros, M.
<b>Journal:</b>	<i>Journal of Applied Econometrics</i>
<b>DOI/Link:</b>	<a href="https://doi.org/10.1002/jae.2512">10.1002/jae.2512</a>
<b>Document version:</b>	Accepted manuscript (post-print)
<b>Document license:</b>	<i>[If the document is published under a Creative Commons, enter link to the license here]</i>

### **General Rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

If the document is published under a Creative Commons license, this applies instead of the general rights.

# MODELING AND FORECASTING LARGE REALIZED COVARIANCE MATRICES AND PORTFOLIO CHOICE

**Laurent A.F. Callot**

VU University Amsterdam, CREATES,  
and the Tinbergen Institute  
E-mail: [l.callot@vu.nl](mailto:l.callot@vu.nl)

**Anders B. Kock<sup>1</sup>**

Aarhus University and CREATES  
E-mail: [akock@creates.au.dk](mailto:akock@creates.au.dk)

**Marcelo C. Medeiros**

Pontifical Catholic University of Rio de Janeiro  
E-mail: [mcm@econ.puc-rio.br](mailto:mcm@econ.puc-rio.br)

**Abstract:** We consider modeling and forecasting large realized covariance matrices by penalized vector autoregressive models. We consider Lasso-type estimators to reduce the dimensionality and provide strong theoretical guarantees on the forecast capability of our procedure. We show that we can forecast realized covariance matrices almost as precisely as if we had known the true driving dynamics of these in advance. We next investigate the sources of these driving dynamics as well as the performance of the proposed models for forecasting the realized covariance matrices of the 30 Dow Jones stocks. We find that the dynamics are not stable as the data are aggregated from the daily to lower frequencies. Furthermore, we are able to beat our benchmark by a wide margin. Finally, we investigate the economic value of our forecasts in a portfolio selection exercise and find that in certain cases an investor is willing to pay a considerable amount in order to get access to our forecasts.

**Keywords:** Realized covariance; shrinkage; Lasso; forecasting; portfolio allocation.

**JEL:** C22.

**Acknowledgements:** We would like to thank two anonymous referees and the editor Jonathan Wright for constructive comments which have greatly improved the paper. Parts of the research for this paper were done while the first and second authors were visiting the Pontifical Catholic University of Rio de Janeiro.

---

<sup>1</sup>Corresponding author: Department of Economics and Business Economics - CREATES, Fuglesangs Allé 4, building 2628, 313, 8210 Aarhus V, Denmark, Phone: +4587165498.

Its hospitality is gratefully appreciated. MCM's research is partially supported by the CNPq/Brazil. Financial support by the Danish National Research Foundation (grant DNR78, CREATES) is gratefully acknowledged.

## 1. INTRODUCTION

This paper deals with modeling and forecasting large-dimensional time-varying realized measures of covariance matrices of returns on financial assets. By realized measures of a covariance matrix we mean estimates of the integrated covariance matrix based on ultra-high-frequency data as, for example, the composite realized kernel of [Lunde et al. \(2015\)](#). We evaluate the proposed model in terms of its forecasting performance in comparison to benchmark alternatives as well as in terms of its economic value in a conditional mean-variance analysis to assess the value of volatility forecasts to short-horizon investors as in [Fleming et al. \(2003\)](#). Our procedure is able to unveil the dynamics governing the evolution of the covariance matrices and we investigate which sectors are instrumental in driving these dynamics. This is important since we can answer questions such as: are certain sectors uncorrelated with others? Are the variances of stocks mainly described by their own past or are there spillover effects? Are the covariances governed by the same dynamics as the variances? We also investigate whether these dynamics change as the realized measures are aggregated from the daily to the weekly and monthly level as one could expect the driving dynamics to differ from the short to the long run.

Modern portfolio selection as well as risk management and empirical asset pricing models strongly rely on precise forecasts of the covariance matrix of the assets involved. For instance, the traditional mean-variance approach of [Markowitz \(1952\)](#) requires the estimation or modeling of all variances and covariances. The evolution of financial markets has increased the number of assets, leading traditional approaches to be less suitable to be used by practitioners. Furthermore, [Bollerslev et al. \(1988\)](#) finds that covariance matrices appear to be quite variable over time with a strong autoregressive structure, and find expected returns to be strongly correlated with these

variations. Typical multivariate ARCH-type models fail to deliver reliable estimates due the curse of dimensionality and large computational burdens. Possible solutions frequently used in practice are: (1) a weighted-average of past squared returns as in the Riskmetrics methodology; (2) vast latent conditional covariance models, or (3) the construction of factor models. In this paper we will take a different route and will consider the estimation of vast vector autoregressive models for realized measures of covariance matrices as in [Kock and Callot \(2015\)](#). To avoid the curse of dimensionality we advocate the use of the Least Absolute Shrinkage and Selection Operator (Lasso) proposed by [Tibshirani \(1996\)](#). An advantage of this approach is that it does not reduce dimensionality by transforming variables, thus keeping the interpretability of the individual variables. Our paper is related to the work of [Audrino and Knaus \(2014\)](#) who applied the Lasso to univariate autoregressive models in order to forecast realized measures of volatility.

The contributions of this paper are as follows. First, we put forward a methodology to model and forecast large time-varying realized covariance matrices with a minimum number of restrictions. Second, our method can shed light on the drivers of the dynamics of these realized covariance matrices as the Lasso also performs variable selection. Third, we derive an upper bound on the forecast error which is valid even in finite samples. Fourth, we show how this bound translates into a bound on the forecast error of the time-varying variance of a portfolio constructed from a large number of assets. Fifth, we use our method to forecast the realized covariance matrices at the daily, weekly, and monthly level of aggregation. Finally, we apply our methodology to the selection of a portfolio with mean-variance preferences.

We use the estimator of [Lunde et al. \(2015\)](#) to estimate the realized covariance matrix. Several other papers have dealt with the estimation of vast covariance matrices, e.g. [Bickel and Levina \(2008\)](#), [Levina et al. \(2008\)](#), [Fan et al. \(2008\)](#), [Wang and Zou \(2010\)](#), [Fan et al. \(2011\)](#), [Fan et al. \(2012\)](#), [Fan et al. \(2012\)](#), [Hautsch et al. \(2012\)](#), [Fan et al. \(2013\)](#). Our goal is to build an econometric methodology which will be used to construct dynamic models to forecast large covariance

matrices estimated elsewhere. Even with proper estimates of covariance matrices, modeling their dynamics and producing reliable forecasts pose major challenges, especially in high dimensions.

In terms of forecasting (realized) covariance matrices, there are some papers related to ours. However, none of them tackles the curse of dimensionality in a general manner. For example, [Bauer and Vorkink \(2011\)](#) propose a multivariate version of the heterogeneous autoregressive (HAR) model. Their approach is based on the log-matrix covariance specification of [Chiu et al. \(1996\)](#). However, their model is only feasible for covariance matrices of low dimension. Furthermore, the parameters of their model are not easily interpretable because of the log-matrix transformation, and thus one can not investigate the main driving forces of the dynamics of the covariance matrix as in our approach. [Chiriac and Voev \(2011\)](#) consider a multivariate ARFIMA model. However, only small covariance matrices can be modeled. [Golosnoy et al. \(2012\)](#) propose a Conditional Autoregressive Wishart (CAW) model. As before, in their application the authors only consider five different assets.

The use of high-frequency data for portfolio selection has been studied by many authors. [Fleming et al. \(2003\)](#) measure the economic gains of using high-frequency data in the context of investment decisions. Their results indicate that the economic value of switching from daily to intradaily returns to estimate daily covariance matrices are substantial. They estimate that a risk-averse investor would be willing to pay 50 to 200 basis points per year to capture the observed gains in portfolio performance. However, their analysis is restricted to small sets of assets. [Hautsch et al. \(2015\)](#) introduce the Multi-Scale Spectral Components (MSSC) model, which is a kind of factor specification, for forecasting covariance matrices and they show that high-frequency data models can translate into better portfolio allocation decisions over longer investment horizons. Although, these authors consider the same problem as we do here, their modeling approach is quite different.

The rest of the paper is organized as follows. Section 2 describes the problem setup, defines notation, and presents the Lasso and some key assumptions. In Section 3 we present theoretical performance guarantees of our procedure. The data set and computational issues are discussed in Section 4, variable selection and forecasting results are presented in Section 5, portfolio selection is discussed in Section 6. Finally, Section 7 concludes the paper. All proofs are deferred to the supplementary appendix.

## 2. METHODOLOGY

In this section we put forward our methodology and present a finite sample upper bound on the forecast error of our procedure. But first we introduce the following notation.

**2.1. Notation.** For any  $x \in \mathbb{R}^m$ ,  $\|x\| = \sqrt{\sum_{i=1}^m x_i^2}$ ,  $\|x\|_{\ell_1} = \sum_{i=1}^m |x_i|$  and  $\|x\|_{\ell_\infty} = \max_{1 \leq i \leq m} |x_i|$  denote  $\ell_2$ ,  $\ell_1$  and  $\ell_\infty$  norms, respectively. For an  $m \times m$  matrix  $A$ ,  $\|A\|_\infty$  denote the maximum absolute entry of  $A$ . For any two real numbers  $a$  and  $b$ ,  $a \vee b = \max(a, b)$  and  $a \wedge b = \min(a, b)$ .

**2.2. The econometric method.** Let  $\Sigma_t$  denote the  $n_T \times n_T$  realized measure of the integrated covariance matrix as of time  $t$ . Note that  $n_T$  is indexed by the sample size  $T$  implying that  $n_T$  may be large compared to  $T$  and hence standard asymptotics, which take  $n_T$  as a fixed number, may not accurately reflect the actual performance in finite samples.

The entries of the conditional covariance matrix are potentially a complicated function of past entries. If we allowed conditioning on the infinite past, each entry of  $\Sigma_t$  would be a function of infinitely many variables. Instead, we assume that  $\Sigma_t$  only depends on the  $p_T$  most recent values of  $\Sigma_t$ , namely  $\Sigma_{t-1}, \dots, \Sigma_{t-p_T}$ . As each of these matrices has  $(n_T + 1)n_T/2$  unique entries we still have that every entry of  $\Sigma_t$  can be a function of  $n_T \frac{n_T+1}{2} p_T$  variables (plus a constant for each entry). In the case of the Dow Jones Industrial Average (DJIA) considered in this paper we have  $n_T = 30$ , and assuming that  $p_T = 10$  lags suffice to describe the dynamics of  $\Sigma_t$ ,

every entry of  $\Sigma_t$  can be a function of up to 4,650 variables (plus a constant). This is the case for all the 465 unique entries of  $\Sigma_t$ <sup>2</sup>. To make this manageable, we shall assume that every entry of  $\Sigma_t$  is a *linear* function of its past, so that  $\Sigma_t$  follows a VAR( $p_T$ ) process. This allows for rich dynamics as every entry of  $\Sigma_t$  can depend on several thousand variables, leading to equations with many more parameters than observations.

Formally, define  $y_t = \text{vech } \Sigma_t$  where  $\text{vech}$  is the half-vectorization operator returning a vector of length  $n_T(n_T + 1)/2$  with the unique entries of  $\Sigma_t$ . We assume that  $y_t$  follows a vector autoregression of order  $p_T$ , i.e.,

$$y_t = \omega^* + \sum_{i=1}^{p_T} \Phi_i^* y_{t-i} + \epsilon_t, \quad t = 1, \dots, T \quad (1)$$

where  $\Phi_i^*$ ,  $i = 1, \dots, p_T$ , are the  $k_T \times k_T$  dimensional parameter matrices pertaining to the coefficients of the lagged left hand side variables with  $k_T = n_T(n_T + 1)/2$ ,  $\omega^*$  is a  $k_T \times 1$  vector of intercepts allowing for non-zero means of all variances and covariances and  $\epsilon_t$  is the zero-mean error term. We assume that we have  $p_T$  initial values  $y_{-p_T+1}, \dots, y_0$ . Henceforth we suppress the dependence of  $n_T, k_T$  and  $p_T$  on  $T$  to simplify notation.

The number of parameters per equation,  $k$ , increases quadratically in  $n$ , as does the number of equations. So, even for  $\Sigma_t$  of a moderate dimension, the number of parameters in (1) may be very large. Hence, standard estimation techniques such as least squares may provide very imprecise parameter estimates or even be infeasible if the number of variables is greater than  $T$ . To circumvent this problem, we use the Least Absolute Shrinkage and Selection Operator (Lasso) of Tibshirani (1996) which is feasible even when the number of parameters to be estimated is (much) larger than  $T$ .

Even though each entry of  $\Sigma_t$  can depend on every entry of the previous  $p$  conditional covariance matrices, it is reasonable to assume that the major stocks play a more important role in describing the dynamics of  $\Sigma_t$ . Furthermore, it is likely

<sup>2</sup>Actually, we shall work with models including up to 20 lags, thus having 9,300 variables per equation.

that the intrasectoral dynamics are stronger than the intersectoral ones. By this we mean that lags of variables from the same sector as the left hand side variable being modeled are more likely to be significant than variables from other sectors. For these reasons the  $\Phi_i^*, i = 1, \dots, p$ , may contain many zeros, i.e. they are sparse matrices. It is exactly in such a setting that we can establish performance guarantees on the Lasso. Thus, the Lasso can be used to unveil and disentangle the potentially complex dynamics of the sequence of realized covariance matrices. If some of the intercepts in  $\omega$  equal zero this will make the model even sparser. However, this is not a requirement for our theory to be valid. It is important to stress that we do not assume the realized covariance matrices themselves to be sparse. The coefficients  $\Phi_i^*$ , which describe the dynamics of the realized covariance matrices, are the quantities assumed to be sparse.

Let  $Z_t = (1, y'_{t-1}, \dots, y'_{t-p})'$  be the  $(kp + 1) \times 1$  vector of explanatory variables and  $Z = (Z_T, \dots, Z_1)'$  the  $T \times (kp + 1)$  matrix of covariates. Let  $y_i = (y_{T,i}, \dots, y_{1,i})'$  be the  $T \times 1$  vector of observations on the  $i$ th variable,  $i = 1, \dots, k$ , and  $\epsilon_i = (\epsilon_{T,i}, \dots, \epsilon_{1,i})'$  the corresponding vector of error terms. Finally,  $\gamma_i^* = (\omega_i^*, \beta_i^{*'})'$  is the  $(kp + 1)$  dimensional vector of true parameters for equation  $i$  which also implicitly depends on  $T$ . Hence, we may write (1) as

$$y_i = Z\gamma_i^* + \epsilon_i, \quad i = 1, \dots, k \quad (2)$$

such that each equation in (1) may be modeled separately. In our application modeling the DJIA, every equation has  $465 \times p$  explanatory variables plus a constant. However, it is likely that many of these variables are irrelevant in explaining the dynamics of  $y_i$ , leading to many entries of  $\beta_i^*$  being equal to zero. For each  $i = 1, \dots, k$  we denote by  $s_i = |\{j : \beta_{i,j}^* \neq 0\}|$  the number of non-zero entries in  $\beta_i^*$ . We shall require below that  $s_i$  does not grow too fast. It is important to stress that it is the number of non-zero elements of  $\beta_i^*$  *per equation* that is the important quantity. We do not make any assumptions on  $\omega^*$  being sparse, none of the  $\omega_i^*$  are assumed to be zero but this is not ruled out either.



**2.3. The Lasso.** The properties of the Lasso have been studied extensively, as for example in Zhao and Yu (2006), Meinshausen and Bühlmann (2006), Bickel et al. (2009), Kock and Callot (2015), and Medeiros and Mendes (2015) to mention just a few. It is known that it only selects the correct model asymptotically under rather restrictive conditions on the dependence structure of the covariates. However, it can still serve as an effective screening device in these cases as it can remove many irrelevant covariates while keeping the relevant ones and estimating the model with high precision. We investigate the properties of the Lasso when applied to each equation separately and we provide finite sample bounds on the system-wise forecasting performance. As we do not assume that any of the intercepts in  $\omega^*$  is sparse we shall employ a version of the Lasso which does not penalize these. To be precise, we estimate  $\gamma_i^*$  in (2) by minimizing the following objective function

$$L(\gamma_i) = \frac{1}{T} \|y_i - Z\gamma_i\|^2 + 2\lambda_T \|\beta_i\|_{\ell_1} \quad (3)$$

with respect to  $\gamma_i$  where  $\lambda_T$  is a sequence to be defined exactly below. Let  $\hat{\gamma}_i = (\hat{\omega}_i, \hat{\beta}_i)'$  denote the minimizer. (3) is basically the least squares objective function plus an extra term penalizing entries of  $\beta_i$  that are different from zero. The intercepts are not penalized.  $\lambda_T$  is often chosen either by cross validation or using an information criterion. The Lasso performs estimation and variable selection simultaneously which is useful in high-dimensional models as the number of specification tests to be carried out after the usual estimation step would be daunting. Furthermore, it is not clear in which order these tests should be carried out and one would necessarily have to resort to *ad hoc* procedures. Finally, the Lasso is a convex minimization problem, resulting in fast estimation and variable selection.

**2.4. Forecasting with the Lasso.** Once the  $\hat{\gamma}_i$  have been obtained for equation  $i$ , forecasting is done as usual in VAR models. The one-step ahead forecast is given by  $\hat{y}_{i,T+1} = \hat{\gamma}_i' Z_T$ . Doing this for all  $i = 1, \dots, k$  results in a forecast  $\hat{\Sigma}_{T+1}$  of the matrix  $\Sigma_{T+1}$ . Forecasts at longer horizons are obtained by iterating forward the estimated VAR model.

## 3. THEORETICAL RESULTS

In this section we provide an upper bound on the maximal forecast error  $\|\widehat{\Sigma}_{T+1} - \Sigma_{T+1}\|_\infty$ . It is exactly such a bound which is required for providing a theoretical upper bound on the errors of the variance forecasts of a portfolio of assets. Let  $w \in \mathbb{R}^n$  denote a set of portfolio weights. The true conditional variance of the portfolio is given by  $\sigma_{T+1}^2 = w' \Sigma_{T+1} w$ , while the forecasted variance is  $\widehat{\sigma}_{T+1}^2 = w' \widehat{\Sigma}_{T+1} w$ . How close are  $\widehat{\sigma}_{T+1}^2$  and  $\sigma_{T+1}^2$  to each other? In the presence of an upper bound on the positions an investor may take in any given asset we can answer this question by providing an upper bound on  $\|\widehat{\Sigma}_{T+1} - \Sigma_{T+1}\|_\infty$ . Thus, we need a precision guarantee for the forecasts of the Lasso. Assume that  $\|w\|_{\ell_1} \leq 1 + c$  for some  $c \geq 0$ . Here  $c$  indicates the amount of short selling that is allowed.  $c = 0$  corresponds to no short selling, while for any  $c > 0$  the maximum amount of short selling possible is  $c/2$ . For any fixed vector of portfolio weights, the next theorem, of which a similar version may be found in [Fan et al. \(2012\)](#), utilizes the short selling constraint to derive an upper bound on the distance between the predicted portfolio variance and the actual portfolio variance.

**Theorem 1.** *Assume that  $\|w\|_{\ell_1} \leq 1 + c$  for some  $c \geq 0$ . Then,*

$$|\widehat{\sigma}_{T+1}^2 - \sigma_{T+1}^2| \leq \|\widehat{\Sigma}_{T+1} - \Sigma_{T+1}\|_\infty (1 + c)^2.$$

Theorem 1 reveals that in the presence of a restriction on the exposure of the portfolio, an upper bound on  $\|\widehat{\Sigma}_{T+1} - \Sigma_{T+1}\|_\infty$  implies an upper bound on the distance between  $\widehat{\sigma}_{T+1}^2$  and  $\sigma_{T+1}^2$ . It is sensible that a short-selling constraint is necessary in order to establish such a bound since otherwise the investor could go infinitely long in the least precisely forecasted stock. Note that the bound in Theorem 1 is generic in the sense that the forecast  $\widehat{\Sigma}_{T+1}$  does not need to come from a specific method. Any matrix  $\widehat{\Sigma}_{T+1}$  can be plugged in.

We shall now give an upper bound on  $\|\widehat{\Sigma}_{T+1} - \Sigma_{T+1}\|_\infty$  based on our VAR approach. Let  $m_i$  and  $\eta_{i,y}^2$  denote the mean and variance of  $y_{t,i}$ , respectively, and  $\eta_{i,\epsilon}^2$

be the variance of the mean zero  $\epsilon_{t,i}$ ,  $1 \leq i \leq k$ . Then define  $m_T = \max_{1 \leq i \leq k} |m_i|$ ,  $\eta_T = \max_{1 \leq i \leq k} (\eta_{i,y} \vee \eta_{i,\epsilon})$ , and let  $\bar{s} = \max(s_1, \dots, s_k)$  denote the maximal number of non-zero coefficients in any equation of the VAR.

**Theorem 2.** *Let  $\lambda_T = \sqrt{8 \ln(1+T)^5 \ln(1+k)^4 \ln(1+p)^2 \ln(k^2 p) K_T^2 / T}$  where  $K_T$  is defined in the appendix and bounded if  $m_T$  and  $\eta_T^2$  are bounded. Under regularity conditions made precise in the appendix, one has for all  $0 < q < 1$*

$$\|\widehat{\Sigma}_{T+1} - \Sigma_{T+1}\|_\infty \leq \sqrt{\tilde{K}_T} \left( \frac{16}{q\kappa^2} \bar{s} \lambda_T + 1 \right) \quad (4)$$

with high probability (the exact probability, as well as the definition of the constant  $\kappa > 0$ , are given in the appendix) where  $\tilde{K}_T = \frac{\eta_T^2 (9m_T^2 + \ln(kp+1) \ln(T))}{\eta_T^2 \wedge \frac{1}{6}}$ .

Theorem 2 gives an upper bound on the forecast error of  $\widehat{\Sigma}_{T+1}$  which is valid even in finite samples. Recall that  $k = n(n+1)/2$ . Thus, for any size  $n$  of the covariance matrix, Theorem 2 gives an upper bound on the forecast error of the Lasso. Clearly, the larger the covariance matrix to be forecasted is ( $k$  large), the larger this upper bound since there are more elements which have to be forecasted. However,  $\tilde{K}_T$  only increases logarithmically in  $k$ . In order to gauge how precise the upper bound in (4) is, it is instructive to compare it to the bound one could have obtained if the true parameters had been known. If the true parameters  $\Phi_i^*$ ,  $i = 1, \dots, p$  were known, one would have the following upper bound:

**Lemma 1.** *Assume that  $\Phi_i^*$ ,  $i = 1, \dots, p$  are known. Then, under regularity conditions made precise in the appendix, with  $\eta_{T,\epsilon} = \max_{1 \leq i \leq k} \eta_{i,\epsilon}$ , one has*

$$\|\widehat{\Sigma}_{T+1} - \Sigma_{T+1}\|_\infty \leq \sqrt{2\eta_{T,\epsilon}^2 \ln(k) \ln(T)}$$

with high probability (the exact one is given in the appendix).

Compared to the bound in Theorem 2 the bound in Lemma 1 has roughly removed the term  $\sqrt{\tilde{K}_T} \bar{s} \lambda_T$  which is the part of the upper bound stemming from the estimation error of the parameters. The last part of the bound in Theorem 2 stems

from the fact that the error terms in the model (1) are unknown such that we must still bound their fluctuations with high probability. Even when the true parameters are known, these unknown error terms are an irremovable source of forecast error. From an asymptotic point of view one will usually requires that  $\bar{s}\lambda_T \rightarrow 0$  since this is needed to make the estimation error tend to zero in probability. In this case the bound in Theorem 2 is almost as good as the one in Lemma 1. Hence, our forecasts are almost as precise as if we had known the true model from the outset even though we do not assume knowledge of which variables are relevant, nor do we assume to know the values of the non-zero parameters.

By combining Theorems 1 and 2 one may achieve the following upper bound.

**Corollary 1.** *Under the assumptions of Theorems 1 and 2 one has that*

$$|\hat{\sigma}_{T+1}^2 - \sigma_{T+1}^2| \leq \sqrt{\tilde{K}_T} \left( \frac{16}{q\kappa^2} \bar{s}\lambda_T + 1 \right) (1+c)^2$$

*with high probability (the exact one is given in the appendix) and  $\tilde{K}_T$  as in Theorem 2.*

Corollary 1 provides a finite sample upper bound on the error of the forecast of the portfolio variance under a short selling constraint. As stressed previously, in the absence of a short selling constraint, corresponding to  $c \rightarrow \infty$ , the upper bound in the above display tends to  $\infty$  as in the worst case one could go infinitely long in the stock whose volatility is forecasted the least precisely and offsetting this with a short position in, say, the stock whose volatility is forecasted the most precisely. Note also that the bound in Corollary 1 is actually valid uniformly over  $\{w \in \mathbb{R}^n : \|w\|_{\ell_1} \leq 1+c\}$ , i.e. all portfolios with a short selling constraint of  $1+c$ . To be precise,

**Corollary 2.** *Under the assumptions of Theorems 1 and 2 one has that*

$$\sup_{w \in \mathbb{R}^n : \|w\|_{\ell_1} \leq 1+c} |\hat{\sigma}_{T+1}^2 - \sigma_{T+1}^2| \leq \sqrt{\tilde{K}_T} \left( \frac{16}{q\kappa^2} \bar{s}\lambda_T + 1 \right) (1+c)^2$$

with high probability (the exact one is given in the appendix) and  $\tilde{K}_T$  as in Theorem 2.

As  $\hat{w} = \arg \min_{w \in \mathbb{R}^n: \|w\|_{\ell_1} \leq 1+c, \sum_{i=1}^n w_i=1} w' \hat{\Sigma}_{T+1} w$  satisfies the conditions of Corollary 2 we have in particular that  $\hat{w}' \hat{\Sigma}_{T+1} \hat{w}$  and  $\hat{w}' \Sigma_{T+1} \hat{w}$  are close to each other. This means that the forecasted portfolio variance is not far from the actually realized portfolio variance when using the weights  $\hat{w}$  that seek to minimize the out of sample forecast error.

**Ensuring positive-definiteness of  $\hat{\Sigma}_{T+1}$ .**  $\hat{\Sigma}_{T+1}$  are in general not positive definite, except for the no-change forecasts. Furthermore, certain matrices are ill-conditioned, which renders the estimation of portfolio weights unstable or even infeasible. To overcome these issues, we follow Hautsch et al. (2012) and apply eigenvalue cleaning to every matrix that has eigenvalues smaller than or equal to 0 or a condition number greater than  $10n_T$ .

Write the spectral decomposition of  $\hat{\Sigma}_{T+h} = \hat{V}'_{T+h} \hat{\Lambda}_{T+h} \hat{V}_{T+h}$  where  $\hat{V}_{T+h}$  is the matrix of eigenvectors and  $\hat{\Lambda}_{T+h}$  a diagonal matrix with the  $n$  eigenvalues  $\hat{\lambda}_{iT+h}$  on its diagonal. Let  $\lambda_{\text{mp}T+h} = \min\{\hat{\lambda}_{iT+h} | \hat{\lambda}_{iT+h} > 0\}$ , replace all the  $\hat{\lambda}_{iT+h} < \lambda_{\text{mp}T+h}$  by  $\lambda_{\text{mp}T+h}$  and define the diagonal matrix  $\tilde{\Lambda}_{T+h}$  with the cleaned eigenvalues on its diagonal. The regularized forecast matrix  $\tilde{\Sigma}_{T+h} = \hat{V}'_{T+h} \tilde{\Lambda}_{T+h} \hat{V}_{T+h}$  is by construction positive definite.

**Log-matrix transform.** An alternative approach relies on properties of the matrix logarithm and the matrix exponential as explained in Chiu et al. (1996) or Bauer and Vorkink (2011). Prior to estimation we transform the data by applying the matrix logarithm, defining  $\Omega_t = \log(\Sigma_t)$ , and estimating the VAR on  $\text{vech} \Omega_t$ . We then construct a forecast  $\hat{\Omega}_{T+h}$  and revert the matrix logarithm by applying the matrix exponential to the forecasted covariance matrices to create our forecast  $\hat{\Sigma}_{T+h} = \exp(\hat{\Omega}_{T+h})$  which is positive definite by construction. Thus, we now assume that it is  $\text{vech} \Omega_t$  which follows a VAR model as in (1). Imposing exactly the same

regularity conditions as in Theorem 2 (but on the VAR for  $\text{vech } \Omega_t$ ) one can establish the following upper bound on the forecast error<sup>3</sup>.

**Theorem 3.** *Let  $\lambda_T = \sqrt{8 \ln(1+T)^5 \ln(1+k)^4 \ln(1+p)^2 \ln(k^2 p) K_T^2 / T}$  where  $K_T$  is defined in the appendix and bounded if  $m_T$  and  $\eta_T^2$  are bounded. Under regularity conditions made precise in the appendix, one has for all  $0 < q < 1$*

$$\begin{aligned} \|\widehat{\Sigma}_{T+1} - \Sigma_{T+1}\|_\infty &= \|e^{\widehat{\Omega}_{T+1}} - e^{\Omega_{T+1}}\|_\infty \\ &\leq \sqrt{\widetilde{K}_T} \left( \frac{16}{q\kappa^2} \bar{s}\lambda_T + 1 \right) e^{n\sqrt{\widetilde{K}_T} \left( 2[\max_{1 \leq i \leq k} \|\gamma_i^*\|_{\ell_1} + 1] + \left[ \frac{16}{q\kappa^2} \bar{s}\lambda_T + 1 \right] \right)} \end{aligned}$$

with high probability (the exact one, as well as the definition of the constant  $\kappa > 0$ , are given in the appendix) where  $\widetilde{K}_T = \frac{\eta_T^2(9m_T^2 + \ln(kp+1)\ln(T))}{\eta_T^2 \wedge \frac{1}{6}}$ . If  $\Phi_i^*$ ,  $i = 1, \dots, p$  are known and  $\eta_{T,\epsilon} = \max_{1 \leq i \leq k} \eta_{i,\epsilon}$ , one has under regularity conditions made precise in the appendix

$$\begin{aligned} \|\widehat{\Sigma}_{T+1} - \Sigma_{T+1}\|_\infty &= \|e^{\widehat{\Omega}_{T+1}} - e^{\Omega_{T+1}}\|_\infty \\ &\leq \sqrt{\eta_{T,\epsilon}^2 \log(k) \ln(T)} e^{n \left( 2\sqrt{\widetilde{K}_T} [\max_{1 \leq i \leq k} \|\gamma_i^*\|_{\ell_1} + 1] + \sqrt{\widetilde{K}_T} \right)} \end{aligned}$$

with high probability (the exact probability is given in the appendix).

The bound in Theorem 3 is very similar to the one in Theorem 2. The only difference is the part  $e^{n\sqrt{\widetilde{K}_T} \left( 2[\max_{1 \leq i \leq k} \|\gamma_i^*\|_{\ell_1} + 1] + \left[ \frac{16}{q\kappa^2} \bar{s}\lambda_T + 1 \right] \right)}$  which is the price we pay for translating a bound on the forecast error of the log-matrix transformed data into a bound on the original data. The size of this extra term is related to the modulus of continuity of the matrix exponential. Note that the dimension  $n$  enters this term in an exponential manner such that the bound may be rather uninformative when  $n$  is large compared to  $T$ . However, it is not surprising that one must pay such a price as the matrix exponential is a steep function meaning that small changes in its argument can lead to large changes in its value: small changes in  $\|\widehat{\Omega}_{T+1} - \Omega_{T+1}\|_\infty$  can still lead to big changes in  $\|\widehat{\Sigma}_{T+1} - \Sigma_{T+1}\|_\infty = \|e^{\widehat{\Omega}_{T+1}} - e^{\Omega_{T+1}}\|_\infty$ . However, we

<sup>3</sup>We stress again that all notation and assumptions in Theorem 3 are exactly as in Theorem 2 except for the fact that the VAR structure is assumed for  $\text{vech } \Omega_t$  instead of  $\text{vech } \Sigma_t$ .

shall see that in practice the forecasts based on the log-matrix transform are rather precise.

A similar intuition applies when comparing Theorem 3 to the result in Lemma 1. The bounds are similar except for the price paid for translating a bound from the log-matrix world back to our original data. Finally, one should mention that from an economic point of view a drawback of the log-matrix transform is that the parameters no longer have any economic interpretation when the VAR is estimated on  $\text{vech } \Omega_t$  instead of  $\text{vech } \Sigma_t$ . If one is interested in forecasting only, this is of course of no importance.

#### 4. DATA AND IMPLEMENTATION

**Data and Cleaning.** The data consists of 30 stocks from the Dow Jones index from 2006 to 2012 with a total of 1474 daily observations<sup>4</sup>. The daily realized covariances are constructed from 5 minutes returns by the method of Lunde et al. (2015). Results pertaining to the weekly and monthly level of aggregation as well as all tables and figures whose references start with *sp*- can be found in the supplementary material. The stocks can be classified in 8 broad categories highlighted in Table *sp*-1.

When we do not use the log-matrix transformation, the data is transformed by taking the logarithm of the variances ensuring that, after an exponential transformation, all variance forecasts are strictly positive. The sample covered by our data set includes 16 out of the 20 largest intraday point swings of the Dow Jones industrial average, triggered by the financial crisis of 2008 as well as flash-crashes in 2010 and 2011. These events lead to many extreme values in the daily realized covariance matrices. In order to mitigate the effect of these extreme observation we perform some light cleaning of the data prior to estimation. To be precise, we flag for censoring every covariance matrix in which more than 25% of the (unique) entries are more than 4 standard errors (of the series corresponding to that entry) away from their sample average up to then. These matrices are replaced by an average of the

<sup>4</sup>We are grateful to Asger Lunde for kindly providing us with the data.

nearest five preceding and following non-flagged matrices. Replacing these extreme entries with averages including only preceding days leads to replacement with relatively small values as extreme events in our data seem to appear suddenly but decay slowly. In our application we found that the results are robust to variations in the data cleaning procedure as only very few observations are censored (see below). Using this cleaning scheme, the flagged matrices are concentrated in October 2008. The flash crashes of the 6<sup>th</sup> of May 2010 and the 9<sup>th</sup> of August 2011 are also flagged. In total, the number of cleaned matrices is 20 which corresponds to a very small fraction (1.3%) of the total sample size of 1474.

All forecasts are computed recursively and all forecast errors are computed based on the de-transformed forecasts. Daily forecasts are computed using a rolling window of 1000 observations leading to 455 forecasts. The first forecast is made on the 6th of February 2010. We forecast with VAR, EWMA, DCC, and random walk (No-Change) models. The estimators for the VARs are either the Lasso, the adaptive Lasso (adaLasso), using the Lasso as initial estimator, or the post Lasso OLS to be explained in more detail in the next section.

**Implementation.** All the computations<sup>5</sup> are carried out using R and the `lassovar` package based on the `glmnet` package which is an implementation of the coordinate descent algorithm of Friedman et al. (2010). When all  $\beta$  are penalized by the same  $\lambda$  (Lasso) the algorithm standardizes the variables and the estimated parameters are adjusted accordingly; variables are not standardized when estimating the adaLasso.

The intercepts are left unpenalized in all estimations. For all forecasts that do not come from the log-matrix transformed data, positive definiteness is ensured by eigenvalue cleaning. The models are estimated equation by equation, the penalty parameter is chosen by minimizing the Bayesian Information Criterion (BIC). The BIC for equation  $i$  and penalty parameter  $\lambda$  is  $BIC_i(\lambda) = T \times \log(\hat{\epsilon}'_{\lambda,i} \hat{\epsilon}_{\lambda,i}) + \sum_{j=1}^{kp} \mathbb{1}(\hat{\beta}_{ij}^\lambda \neq 0) \log(T)$ , where  $\hat{\epsilon}_{\lambda,i}$  is the estimated vector of error terms corresponding to a penalty of  $\lambda$ . The results of the Lasso are compared to the ones of the

<sup>5</sup>This work was in part carried out on the Dutch national e-infrastructure with the support of SURF Cooperative. Partial replication material can be found at <https://github.com/lcallot/rcv-fc>



adaLasso. Let  $J(\widehat{\beta}_i) = \{j \in \mathbb{R}^{kp} : \widehat{\beta}_{i,j} \neq 0\}$  denote the indices of the coefficients in the  $i$ 'th equation deemed zero by the Lasso and  $\tilde{J}_i(\widehat{\gamma}_i) = \{1\} \cup (J_i(\widehat{\beta}_i) + 1)$  where the set addition is understood elementwise. The adaLasso estimates  $\beta_i^*$  by minimizing the following objective function

$$\tilde{L}(\beta_i) = \frac{1}{T} \left\| y_i - Z_{\tilde{J}_i(\widehat{\gamma}_i)} \gamma_{i, \tilde{J}_i(\widehat{\gamma}_i)} \right\|^2 + 2\lambda_T \sum_{j \in J(\widehat{\beta}_i)} \frac{|\beta_{i,j}|}{|\widehat{\beta}_{i,j}|}, \quad i = 1, \dots, k. \quad (5)$$

If the first stage Lasso estimator classifies a parameter as zero it is not included in the second step, resulting in a problem of a much smaller size. If  $\beta_{i,j}^* = 0$  then  $\widehat{\beta}_{i,j}$  is likely to be small by Theorem 1 in the supplementary appendix and consistency of the Lasso. Hence,  $1/\widehat{\beta}_{i,j}$  is large and the penalty on  $\beta_{i,j}$  is large. Conversely, if  $\beta_{i,j}^* \neq 0$ ,  $\widehat{\beta}_{i,j}$  is not too close to zero and the penalty is small.

We also consider post Lasso least squares estimation ([Belloni and Chernozhukov \(2013\)](#)), that is, estimating the model by OLS after using the Lasso for variable selection.

## 5. DRIVING DYNAMICS AND FORECASTS

This section reports our empirical findings. We begin by studying the dynamic behavior of the sequence of realized covariance matrices by investigating which variables drive its dynamics.

**5.1. Driving dynamics of the realized covariance matrix.** The results in this section are based on a VAR(1) estimated by the Lasso, each of the 465 equations to be estimated has 465 variables (plus a constant). [Table 1](#) reports the average (across estimation windows) fraction of variables from a given category (in rows) selected in equations for stocks belonging to a given category (in columns) at the daily frequency. As an example, 75% of the lagged variances from the Basic Materials sector are included when modeling variances of the Basic Materials sector while only 17% of the lagged variances of stocks belonging to the Consumer, Non-cyclical sector are included when modeling variances of the Basic Materials sector. The numbers are broken down into variance and covariance terms in order to see whether diagonal and

off-diagonal terms have different driving dynamics. There are 30 variance equations and 435 covariance equations. Assigning categories to covariance terms involving variables from two different sectors requires a subjective choice. We have chosen to assign a covariance between two variables from *different* sectors to both sectors. Hence, such intersectoral covariance terms enter as explanatory variables in two rows in Table 1.

The selected models in Table 1 for the variance terms are quite sparse, and lags of variance terms are much more important than lags of covariance terms. The most important dynamics for 7 out of 8 sectors are the intrasectoral ones (own lags). The Industrial and Energy sectors are particularly dependent on their own past as they always use all lagged variances from their respective categories to explain themselves. The financial sector is also driven strongly by its own past. Figure [sp-1](#) in the appendix shows how often each individual stock was selected at each point in time.

The bottom panel of Table 1 investigates the driving forces of the covariance terms. For 6 out of 8 sectors it is lagged variances of that sector which are chosen most often in order to explain the dynamics of the corresponding sector. Off-diagonal terms are now chosen more often than above. However, the fraction of times the off-diagonal terms are chosen is still lower than the corresponding number for the diagonal terms as there are many more off-diagonal terms.

Table [sp-2](#) in the supplementary appendix confirms that the above findings remain valid when including the variance of the S&P 500 index as a common factor.

To complement the selection frequencies in Table 1 and to give a sense of the total number of parameters selected per equation, Figure 1 plots the average number of variables selected in the variance as well as the covariance equations over time for the VAR(1) and VAR(20) estimated by the Lasso. The dimensions of the diagonal equations are remarkably stable and sparse, only including between 11 and 13 variables out of a possible 465 for the VAR(1) and 20 to 24 out of 9300 for the VAR(20).

The off-diagonal equations are less sparse and their dimension, drops around two of the periods of high volatility in May 2010 and August 2011.

Figure 2 displays the percentage of parameters changing classification, that is, the percentage of parameters going from being zero to non-zero or from being non-zero to zero, in two consecutive models. We report results for a VAR(1) and a VAR(20) estimated with the Lasso and for a VAR(1) estimated by Lasso where the data has been transformed using the log-matrix function prior to estimation. For both VAR(1) models the number of changes in the variance (diagonal) equations is around 0.1% while the corresponding number for the VAR(20) is around 0.02%. This leads to 0.5 (VAR(1)) and 2 (VAR(20)) parameters changing from zero to non-zero or vice versa on average. The covariance equations are less stable except when the data has been transformed using the log-matrix function prior to estimation.

In the left panels of Figure 3 we investigate how the  $\ell_2$  forecast error of the one-step ahead forecasts evolves throughout the sample, and in particular how it reacts to the periods of extreme volatility. The forecast errors of the covariances react stronger to the periods of extreme volatility than the forecast errors of the variances. The right panels of Figure 3 shows that the decrease in the model size of the covariance equations is associated with an increase in the penalty parameter  $\lambda$  which determines the amount of shrinkage.

Tables [sp-3](#) and [sp-5](#) in the supplementary appendix report the selection frequencies at the weekly and monthly levels of aggregation, respectively. The diagonal pattern for the explanatory variables of the variances disappears almost entirely at the monthly frequency. The dynamics of the variance equations are now governed mainly by the financial sector and the energy sector, which we interpret as implying that these sectors are driving the long run volatility of the Dow Jones index.

To shed even further light on the findings in Table 1 we also consider the variable selection pattern for 4 individual stocks in more detail. Figure ?? contains the results for Alcoa, IBM, JPM Morgan, and Coca-Cola which belong to the Basic Materials, Technology, Financial and Consumer (non-cyclical) sectors, respectively. For each of

these 4 stocks, we indicate in the corresponding plot whether the lagged variance of the stock indicated on the y-axis is selected in the model estimated at the date given on the x-axis. The variable selection pattern is stable over time, with many lagged variances being either selected or left out throughout the entire sample. A common feature for the four stocks is that a relatively sparse set of variables describes their dynamics and that lags of the stock being modeled are often important.

**5.2. Forecasts.** In this section we forecast the daily realized covariances using the proposed procedures. Results for the weekly and monthly levels of aggregation can be found in the supplementary material. Let  $\widehat{e}_{T+h}$  denote  $\text{vech}(\widehat{\Sigma}_{T+h} - \Sigma_{T+h})$  where  $\widehat{\Sigma}_{T+h}$  denotes the forecast of  $\Sigma_{T+h}$  as of time  $T$  at horizon  $h \geq 1$ . Thus,  $\widehat{e}_{T+h}$  is the vector of forecast errors at horizon  $h$ . Let the times of the first and last forecast be denoted by  $T_1$  and  $T_n$ , respectively. Then, in order to gauge the precision of the procedures above we consider the following measures.

- (1) Average  $\ell_2$ -forecast error ( $\ell_2$ ):  $\frac{1}{T_n - T_1 + 1} \sum_{T=T_1}^{T_n} \|\widehat{e}_{T+h}\|$ . We shall employ this measure separately for the whole matrix, the variances, and the covariances. When it is employed for the whole matrix it corresponds to the average forecast error in the Frobenius norm.
- (2) Average median absolute forecast error (AMedAFE):  $\frac{1}{T_n - T_1 + 1} \sum_{T=T_1}^{T_n} \text{median}(|\widehat{e}_{T+h}|)$ . This measure is considered for the whole matrix, the variances, and the covariances. It is more robust to outliers than the  $\ell_2$ -norm.
- (3) Average maximal absolute forecast error (AMaxAFE):  $\frac{1}{T_n - T_1 + 1} \sum_{T=T_1}^{T_n} \max(|\widehat{e}_{T+h}|)$ . This measure is considered for the whole matrix, the variances, and the covariances. As the maximal forecast error plays a crucial role in Theorems 1 and 2 it is of interest to investigate this quantity for different procedures.

The principal benchmark against which we gauge the VAR forecasts are no-change forecasts (random walk,  $\widehat{\Sigma}_{T+h} = \Sigma_T$  for  $h \geq 1$ .) Table 2 reports the value of the above measures for the no-change forecasts on a gray background. The numbers for all other models are the relative forecast errors compared to the no-change forecast.

Thus, numbers less than one indicate more precise forecasts than the no-change forecasts.

Table 2 also reports results for two alternative models for comparison purposes.

- (1) DCC(1,1) 1-step ahead forecasts for the daily and weekly data, the training sample is too short to estimate the DCC model at the monthly level of aggregation.
- (2) Exponentially Weighted Moving Average (EWMA) with smoothing parameter  $\lambda = 0.96$ . We report 1-step ahead EWMA forecasts at every level of aggregation.

To assess the effect of filtering the data, we report daily no-change forecasts for the uncensored series relative to the censored series in Table 2. The forecast errors are larger for the uncensored data in particular on the AMaxAFE measure, yet the AMaxAFE is even larger for both DCC and EWMA than for the no-change forecasts on uncensored data. Table 2 also shows that the eigenvalue cleaning procedure used to regularize the VAR forecasts (when not using the log-matrix transformation) has a negligible effect on the accuracy of the forecasts.

Variances appear to be harder to forecast than covariances as can be seen from the larger median and maximal average forecast errors for the variance equations. The gains in forecasting accuracy from using VARs estimated by the Lasso are largest for the variance series. We believe this is an important and new finding.

Forecasts from VAR(1) models estimated using the Lasso consistently deliver lower maximal errors than the benchmark. The improvements are more marginal in terms of median and  $\ell_2$  norm of the errors, however small improvements in the precision of the forecasts can result in large gains as the portfolio section will demonstrate.

The post-Lasso appears to provide the most accurate model for  $h = 1$  of the approaches not based on the log-matrix transform and thus it shall be our main focus. At every forecast horizons the post Lasso has a lower median forecast error for the variances than the no-change forecasts, but it is less precise when it comes

to forecasting the covariances. The forecasts from the VAR(1) based on the log-matrix approach are as precise as the ones from the post-Lasso for the non-diagonal elements and superior to these as well as to the no-change forecasts for the covariances. The overall median forecast error should be interpreted with care as it mainly reflects the median forecast error of the off-diagonal terms of the covariance matrix. When considering the maximal forecast error, the VAR(1) estimated by the post-Lasso greatly outperforms the no-change forecasts for the diagonal terms. In fact, it is around 20 percent more precise when forecasting 20 days ahead. The no-change forecasts are still more precise when it comes to the forecast precision of the covariances. More importantly, the VAR(1) based on the log-matrix transform is superior to the no-change forecasts for variances as well as covariances as it is much more precise than the post Lasso approach for the covariances. The inclusion of the S&P500 in the model does not appear to consistently improve the forecasts which is in line with the findings of the previous section where the S&P500 was shown not to have a particularly strong explanatory power compared to the other variances.

Figure 1 showed that more variables were selected in models for the covariance terms than in those for the variances in a VAR(1). For that reason, in order to better capture the dynamics of the off-diagonal terms, we also experimented with gradually including more lags. Results for VAR(20) models in Table 2 show substantial improvement relative to the benchmark for every model and horizon. Specifically, the increased number of lags renders the post Lasso superior to the no-change forecasts for variances as well as covariances. The forecasts of the covariances improve markedly compared to the VAR(1) model. The VAR(20) on the log-matrix transformed data still delivers more precise forecasts for the covariance terms when considering the maximal forecast errors. These forecasts are also much more precise than the no-change forecasts. The lowest  $\ell_2$  forecast errors are also found for the VAR(20) based on the log-matrix transform.

Finally, Tables [sp-4](#) and [sp-6](#) in the supplementary appendix confirm that the variances are harder to forecast than the covariances, even at the weekly and monthly levels of aggregation.

In conclusion, the forecast precision of the VAR models is in general improved by including more lags hinting at the presence of serial dependence, in particular at the daily level of aggregation. The particular version of the Lasso used to estimate the VAR model does not seem overly important, which makes the procedure rather robust. The log-matrix approach performs well overall. Finally, the greatest gains in forecast precision are obtained at the longest forecast horizons irrespective of the level of aggregation of the data.

## 6. PORTFOLIO SELECTION

In this section we use the forecasted covariance matrices to construct investment portfolios. Our analysis is based on [Fleming et al. \(2003\)](#). We consider a risk-averse investor who uses conditional mean-variance analysis to allocate resources across the  $n = 30$  different firms that compose the Dow Jones index. The investor's utility at each point in time is given by

$$U(r_{pt}) = (1 + r_{pt}) - \frac{\gamma}{2(1 + \gamma)} (1 + r_{pt})^2, \quad (6)$$

where  $r_{pt}$  is the portfolio return at time  $t$  and  $\gamma$  is the investor's risk aversion coefficient. The larger the value of  $\gamma$ , the more risk averse is the investor.

Although we have estimated portfolios for daily, weekly and monthly frequencies we will focus on the daily results. The portfolios are rebalanced in each period as follows. Let  $r_{t+1}$ ,  $\hat{\mu}_{t+1}$ , and  $\hat{\Sigma}_{t+1}$  denote, respectively, an  $n \times 1$  vector of stock returns at time  $t + 1$ , the expected return at time  $t + 1$  given data up to time  $t$  and the forecasted conditional covariance matrix at time  $t + 1$  based on information up to period  $t$ .  $\hat{\Sigma}_{t+1}$  is computed by the methods described earlier while  $\hat{\mu}_{t+1}$  is computed by a moving average of 100 days.<sup>6</sup>

<sup>6</sup>We have experimented with different lengths of the moving average window and the results did not change significantly.

The investor's problem at  $t = t_0, \dots, T - 1$  is to select a vector of weights for period  $t + 1$  based solely on information up to time  $t$ . The investor chooses the weights that minimize the portfolio volatility, subject to a target expected return and several weight constraints:

$$\begin{aligned} \hat{w}_{t+1} &= \arg \min_{w_{t+1}} w'_{t+1} \hat{\Sigma}_{t+1} w_{t+1} \\ \text{s.t.} \quad & w'_{t+1} \hat{\mu}_{t+1} = \mu_{\text{target}}, \quad \sum_{i=1}^n w_{it+1} = 1, \\ & \sum_{i=1}^n |w_{it+1}| \mathbb{I}(w_{it} < 0) \leq 0.30 \quad \text{and} \quad |w_{it+1}| \leq 0.20, \end{aligned} \quad (7)$$

where  $w_{t+1}$  is an  $n \times 1$  vector of portfolio weights on the stocks,  $\mu_{\text{target}}$  is the target expected rate of return from  $t$  to  $t + 1$  and  $\mathbb{I}(\cdot)$  is an indicator function.

The optimal weights  $\hat{w}_{t+1}$  is a function of  $\mu_{\text{target}}$ ,  $\hat{\mu}_{t+1}$ , and  $\hat{\Sigma}_{t+1}$ . We impose two additional restrictions. First, we allow the maximum leverage to be 30% (corresponding to  $c = 0.6$  in Section 3). This yields much more stable portfolios. Second, we restrict the maximum weights on individual stocks to be 20%. This imposes diversification in the portfolio. The target expected return is set to 10% per year. As all the stocks considered are very liquid, we set the transactions cost to be 0.1%. Finally, note that we focus on portfolio optimization based on the one-step ahead forecasts even though the Lasso based techniques actually had their biggest advantages in the long-horizon forecasts. We do so since it is likely that the investor will rebalance the portfolio at the same frequency he has chosen to aggregate the data to.

**Results.** Overall, eight different forecasting models have been considered: five Lasso-based VARs and benchmarks based on the DCC, the EWMA and no-change forecasts. The following statistics are considered.

- (1) Max weight:  $\max_{t_0+1 \leq t \leq T} \max_{1 \leq i \leq n} (\hat{w}_{it})$  for  $t = t_0 + 1, \dots, T$  and  $i = 1, \dots, N$ .

This is the maximum weight over all assets and all time periods.



- (2) Min weight:  $\min_{t_0+1 \leq t \leq T} \min_{1 \leq i \leq n} (\hat{w}_{it})$  for  $t = t_0 + 1, \dots, T$  and  $i = 1, \dots, n$ .

This is the minimum weight over all assets and all time periods.

- (3) Proportion of leverage:  $\frac{1}{n(T-t_0)} \sum_{t=t_0+1}^T \sum_{i=1}^n \mathbb{1}(\hat{w}_{it} < 0)$ . The proportion of leverage is the fraction of negative weights computed for all assets and all time periods.

- (4) Average turnover:  $\frac{1}{n(T-t_0)} \sum_{t=t_0+1}^T \sum_{i=1}^n |\hat{w}_{it} - \hat{w}_{it}^{\text{hold}}|$ , where  $\hat{w}_{it}^{\text{hold}} = \hat{w}_{it-1} \frac{(1+r_{it-1})}{1+r_{pt-1}}$ .

The turnover measures the average change in the portfolio weights.  $w_{it}^{\text{hold}}$ ,  $i = 1, \dots, n$ , are the weights of the hold portfolio. The hold portfolio at period  $t + 1$  is defined as the portfolio resulted from keeping the stocks from period  $t$ .

- (5) Average return:  $\mu_p = \frac{1}{(T-t_0)} \sum_{t=t_0+1}^T r_{pt} = \frac{1}{(T-t_0)} \sum_{t=t_0+1}^T \hat{w}'_t r_t$ . This is the out-of-sample average portfolio return.

- (6) Accumulated return:  $\prod_{t=t_0+1}^T (1 + r_{pt})$ . This is the accumulated portfolio return over the out-of-sample period.

- (7) Standard deviation:  $\sigma_p = \sqrt{\frac{1}{(T-t_0)} \sum_{t=t_0+1}^T \left( r_{pt} - \frac{1}{(T-t_0)} \sum_{t=t_0+1}^T r_{pt} \right)^2}$ . This is the portfolio return standard deviation over the out-of-sample period.

- (8) Sharpe ratio:  $\frac{\mu_p}{\sigma_p}$ . The larger the Sharpe ratio, the better the portfolio as it delivers higher ratios of return over risk.

- (9) Average diversification ratio:  $\frac{1}{(T-t_0)} \sum_{t=t_0+1}^T \frac{\sum_{i=1}^n \hat{w}_{it} \sigma_{it}}{\sigma_{pt}}$ , where  $\sigma_{pt} = \hat{w}'_t \Sigma_t \hat{w}'_t$ . This is the average ratio of weighted standard deviations of the individual stocks over portfolio standard deviations.

- (10) Economic value: the economic value is the value of  $\Delta$  such that, for different portfolios  $p_1$  and  $p_2$ , we have  $\sum_{t=t_0+1}^T U(r_{p_1 t}) = \sum_{t=t_0+1}^T U(r_{p_2 t} - \Delta)$ . It represents the maximum return the investor would be willing to sacrifice each period in order to get the performance gains associated with switching to the second portfolio. In this comparison our benchmark,  $p_1$ , will always be no-change forecasts. We report the value of  $\Delta$  as an annualized basis point fee.

Table 3 contains the results. The models estimated with the log-matrix transformation are superior along most performance criteria. This is in line with the results in Section 5.2. Focusing on the log-matrix transformed data, the VAR(20) is a slightly superior to the VAR(1). The former has average returns of  $6.27 \times 10^{-2}\%$  while the corresponding number for the latter is  $5.89 \times 10^{-2}\%$ . The accumulated returns are 30.02% and 27.77%, respectively, over the 455 days considered. The best model among those not using log-matrix transformed data is the VAR(1) estimated by adaLasso, which delivers an average return of  $3.58 \times 10^{-2}\%$  and an accumulated return equal to 15.16%.

Turning to the Sharpe-ratios, the log-matrix based models are again superior. The Sharpe ratios are  $6.29 \times 10^{-2}$  for the VAR(20) and  $5.87 \times 10^{-2}$  for the VAR(1) which should be compared to a ratio of  $3.65 \times 10^{-2}$  for the best model not based on the log-matrix transform. The standard deviations do not vary much. A possible reason for this is the imposition of the leverage and maximum weight constraints, restricting the exposure of the portfolio.

We turn to models which do not rely on the log-matrix transform. The DCC and EWMA models perform poorly as expected given their inferior forecasts. In terms of accumulated returns, the Lasso based models deliver higher figures than the no-change forecasts. The VAR(1) adaLasso is the model which delivers the highest accumulated returns, 15.16%. It is followed by the VAR(20) estimated by the plain Lasso, 11.45%. The benchmark alternative has an accumulated return of 7.13% while DCC and EWMA have negative accumulated returns. These are big differences, especially when considering that the annual target return over approximately 252 trading days is 10%. The VAR(1) adaLasso model also has the highest Sharpe ratio among the non-log-matrix transformed models. In terms of diversification leverage, all models perform in a very similar way. However, the turnover of some the Lasso based models is lower than the corresponding turnover for the benchmark portfolios.

Finally, we report the annualized economic value of the Lasso over the no-change, DCC and EWMA models. All models deliver a positive annualized economic value

with a maximum of 16.7% over the EWMA. The poor quality of the forecasts based on EWMA and DCC render these models unsuited for portfolio selection. We consider that the relevant economic value benchmark for our procedure is the no-change model. Even in this case the economic value is above 10% for every choice of  $\gamma$ .

## 7. CONCLUSIONS

We considered modeling and forecasting of vast realized covariance matrices. Our approach used the Lasso to reduce the vast dimensionality to a manageable one. We established upper bounds on the forecast error of our procedure which almost coincide with the one obtainable from an oracle procedure thus providing strong theoretical performance guarantees. Next, we investigated the driving dynamics of the covariance matrices. We found that at the short horizon the dynamics were mainly intrasectoral while at the longer horizons the dynamics were often described by the energy and financial sector. A forecast study revealed that our procedure outperforms the benchmark. Finally, we use our covariance matrix forecast in a portfolio selection problem in order to assess the economic value of these forecasts. At the daily level, the Lasso based forecasts have an economic value of up to 16.7% per year.

## REFERENCES

- Audrino, F. and S. Knaus (2014). Lassoing the HAR model: A model selection perspective on realized volatility dynamics. *Econometric Reviews*.
- Bauer, G. and K. Vorkink (2011). Forecasting multivariate realized stock market volatility. *Journal of Econometrics* 2, 93–101.
- Belloni, A. and V. Chernozhukov (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli* 19(2), 521–547.
- Bickel, P. and E. Levina (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics* 36, 199–227.
- Bickel, P., Y. Ritov, and A. Tsybakov (2009). Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics* 37, 1705–1732.

- Bollerslev, T., R. Engle, and J. Wooldridge (1988). A capital asset pricing model with time-varying covariances. *The Journal of Political Economy* 96, 116–131.
- Chiriac, R. and V. Voev (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics* 26, 893–1057.
- Chiu, T., T. Leonard, and K.-W. Tsui (1996). The matrix-logarithmic covariance model. *Journal of the American Statistical Association* 91, 198–210.
- Fan, J., Y. Fan, and J. Lv (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* 147, 186–197.
- Fan, J., Y. Li, and Y. Ke (2012). Vast volatility matrix estimation using high frequency data for portfolio selection. *Journal of American Statistical Association* 107, 412–428.
- Fan, J., Y. Liao, and M. Mincheva (2011). High dimensional covariance matrix estimation in approximate factor models. *The Annals of Statistics* 39, 3320–3356.
- Fan, J., Y. Liao, and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements (with discussion). *Journal of Royal Statistical Society. Series B* 75, 603–680.
- Fan, J., J. Zhang, and K. Yu (2012). Vast portfolio selection with gross-exposure constraints. *Journal of American Statistical Association* 107, 592–606.
- Fleming, J., C. Kirby, and B. Ostdiek (2003). The economic value of volatility timing using realized volatility. *Journal of Financial Economics* 67, 473–509.
- Friedman, J., T. Hastie, and R. Tibishirani (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33, 1–22.
- Golosnoy, V., B. Gribisch, and R. Liesenfeld (2012). The conditional autoregressive wishart model for multivariate stock market volatility. *Journal of Econometrics* 167, 211–223.
- Hautsch, N., L. Kyj, and P. Malec (2015). Do high-frequency data improve high-dimensional portfolio allocations? *Journal of Applied Econometrics* 30, 263–290.

- Hautsch, N., L. Kyj, and R. Oomen (2012). A blocking and regularization approach to high-dimensional realized covariance estimation. *Journal of Applied Econometrics* 27, 625–645.
- Kock, A. and L. Callot (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics* 186, 325–344.
- Levina, E., A. Rothman, and J. Zhu (2008). Sparse estimation of large covariance matrices via a nested Lasso penalty. *The Annals of Applied Statistics* 2, 245–263.
- Lunde, A., N. Shephard, and K. Sheppard (2015). Econometric analysis of vast covariance matrices using composite realized kernels. *Journal of Business and Economic Statistics*. Forthcoming.
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance* 7, 77–91.
- Medeiros, M. and E. Mendes (2015).  $\ell_1$ -regularization of high-dimensional time-series models with non-gaussian and heteroskedastic innovations. *Journal of Econometrics*. forthcoming.
- Meinshausen, N. and P. Bühlmann (2006). High-dimensional graphs and variable selection with the Lasso. *The Annals of Statistics* 34, 1436–1462.
- Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society. Series B (Methodological)* 58, 267–288.
- Wang, Y. and J. Zou (2010). Vast volatility matrix estimation for high-frequency financial data. *The Annals of Statistics* 38, 943–978.
- Zhao, P. and B. Yu (2006). On model selection consistency of Lasso. *The Journal of Machine Learning Research* 7, 2541–2563.

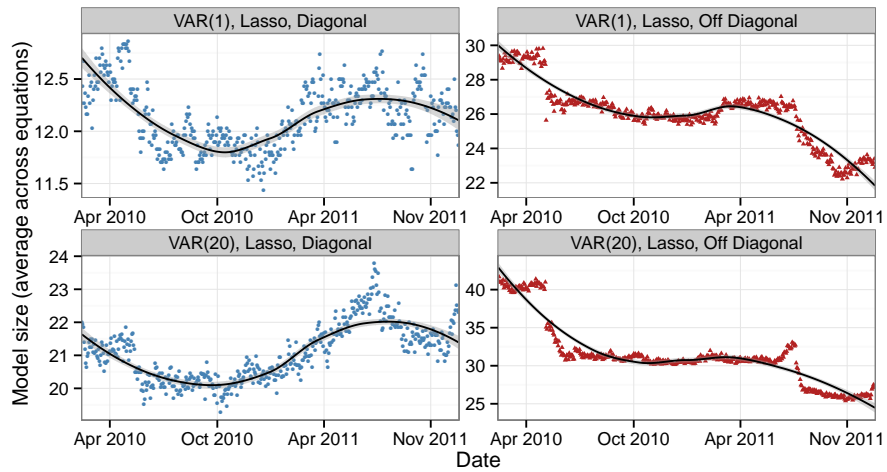


FIGURE 1. Average Equation size of the VAR(1) and VAR(20) estimated by the Lasso.

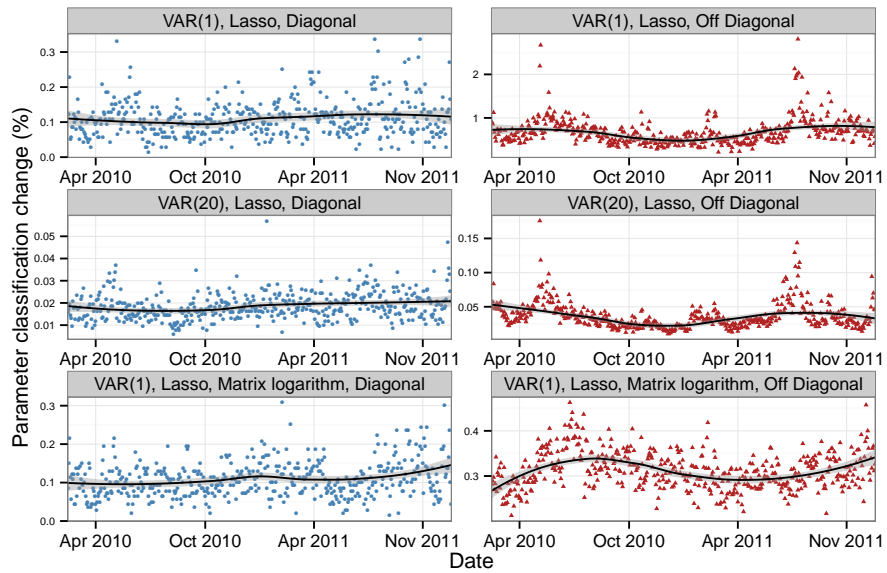


FIGURE 2. Parameter stability: Fraction of parameters that change from being zero to non-zero or vice versa in two consecutive periods.

		Variance Equations							
Lagged variance	Basic Materials	0.75	0.40	0.14	0.52	0.23	0.35	0.57	0.39
	Consumer, Non-cyclical	0.17	0.48	0.37	0.37	0.24	0.20	0.26	0.32
	Financial	0.00	0.42	0.99	0.24	0.64	0.20	0.12	0.48
	Communications	0.32	0.23	0.10	0.57	0.19	0.14	0.27	0.19
	Industrial	0.00	0.19	0.28	0.16	1.00	0.08	0.07	0.18
	Energy	0.58	0.45	0.46	0.33	0.02	1.00	0.38	0.55
	Technology	0.34	0.19	0.09	0.24	0.02	0.05	0.63	0.12
	Consumer, Cyclical	0.34	0.54	0.35	0.29	0.30	0.20	0.31	0.70
Lagged covariance	Basic Materials	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.01
	Consumer, Non-cyclical	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	Financial	0.00	0.01	0.02	0.00	0.00	0.00	0.01	0.00
	Communications	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Industrial	0.02	0.01	0.01	0.02	0.03	0.00	0.03	0.02
	Energy	0.01	0.03	0.01	0.03	0.01	0.01	0.02	0.03
	Technology	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00
	Consumer, Cyclical	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		Covariance Equations							
Lagged variance	Basic Materials	0.81	0.27	0.10	0.24	0.24	0.93	0.26	0.29
	Consumer, Non-cyclical	0.48	0.71	0.56	0.32	0.28	0.35	0.36	0.41
	Financial	0.13	0.25	0.64	0.16	0.06	0.32	0.15	0.17
	Communications	0.62	0.57	0.54	0.65	0.51	0.70	0.61	0.58
	Industrial	0.13	0.13	0.18	0.18	0.34	0.08	0.10	0.07
	Energy	0.12	0.08	0.21	0.06	0.03	0.56	0.11	0.12
	Technology	0.74	0.49	0.51	0.52	0.43	0.34	0.82	0.53
	Consumer, Cyclical	0.14	0.52	0.55	0.37	0.37	0.51	0.29	0.90
Lagged covariance	Basic Materials	0.09	0.12	0.12	0.11	0.09	0.12	0.09	0.11
	Consumer, Non-cyclical	0.04	0.05	0.05	0.04	0.03	0.05	0.04	0.03
	Financial	0.11	0.14	0.18	0.11	0.07	0.10	0.10	0.10
	Communications	0.07	0.09	0.09	0.11	0.04	0.10	0.08	0.08
	Industrial	0.17	0.14	0.15	0.14	0.24	0.11	0.14	0.15
	Energy	0.16	0.16	0.16	0.15	0.10	0.24	0.15	0.16
	Technology	0.08	0.07	0.09	0.08	0.05	0.07	0.08	0.06
	Consumer, Cyclical	0.05	0.05	0.05	0.06	0.04	0.08	0.05	0.06

TABLE 1. Fraction of variables selected from each sector (in row) when modeling variables from the sector in the columns. Model: VAR(1), Lasso, daily.

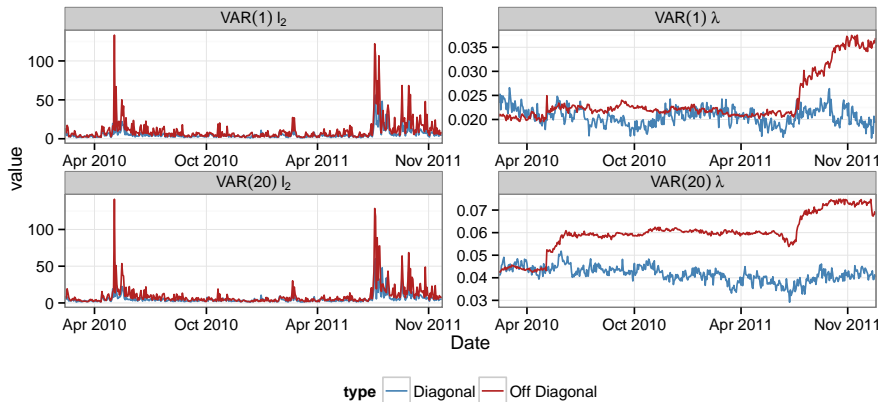
Model	h	AMedAFE			AMaxAFE			$\ell_2$		
		A	D	O	A	D	O	A	D	O
No-Change	1	0.33	0.57	0.33	3.53	3.53	1.47	11.22	5.98	9.22
Censored	5	0.46	0.79	0.45	4.51	4.51	1.91	15.02	7.89	12.41
Benchmark	20	0.58	0.98	0.57	5.12	5.12	2.22	18.05	9.25	15.17
No-Change	1	1.09	1.14	1.06	1.73	1.73	1.59	1.33	1.57	1.17
Uncensored	5	1.20	1.22	1.18	1.60	1.60	1.50	1.32	1.49	1.22
	20	1.14	1.18	1.14	1.53	1.53	1.43	1.27	1.42	1.18
DCC	1	1.70	1.67	1.67	2.38	2.37	2.91	1.99	2.07	1.97
EWMA( $\lambda = 0.96$ )	1	2.67	1.89	2.67	2.29	2.27	3.10	2.57	2.10	2.80
VAR(1), Lasso	1	1.12	1.07	1.12	0.95	0.94	1.17	1.07	0.99	1.11
	5	0.96	0.92	0.96	0.84	0.81	1.18	0.95	0.86	0.99
	20	1.19	0.98	1.19	0.85	0.79	1.42	1.11	0.88	1.19
VAR(1), Lasso	1	1.03	0.96	1.00	0.87	0.86	1.20	1.00	0.90	1.05
Post Lasso OLS	5	0.98	0.92	0.98	0.84	0.82	1.17	0.96	0.87	1.00
	20	1.05	0.95	1.05	0.85	0.80	1.32	1.03	0.87	1.08
VAR(1), adaptive Lasso	1	1.12	1.09	1.12	0.98	0.97	1.23	1.09	1.02	1.13
Initial estimator: Lasso	5	0.96	0.94	0.98	0.86	0.84	1.21	0.96	0.88	1.01
	20	1.07	1.00	1.07	0.87	0.82	1.41	1.08	0.91	1.14
VAR(1), Lasso	1	1.06	1.02	1.06	0.92	0.92	0.97	1.01	0.96	1.03
Log-matrix transform	5	0.91	0.92	0.91	0.79	0.79	0.85	0.88	0.84	0.90
	20	0.83	0.96	0.82	0.79	0.79	0.82	0.85	0.87	0.83
VAR(1), Lasso	1	1.15	1.07	1.12	0.95	0.94	1.18	1.07	0.99	1.11
Not eigenvalue cleaned	5	0.98	0.92	0.98	0.85	0.80	1.21	0.96	0.86	1.00
	20	1.19	0.97	1.19	0.89	0.78	1.50	1.11	0.87	1.19
VAR(1), Lasso	1	1.12	1.07	1.12	0.95	0.94	1.20	1.11	0.99	1.11
Including SP 500	5	0.96	0.94	0.96	0.84	0.81	1.22	0.98	0.87	0.99
	20	1.17	0.97	1.18	0.85	0.79	1.42	1.13	0.87	1.18
VAR(20), Lasso	1	1.06	1.00	1.06	0.90	0.90	1.10	1.01	0.93	1.05
	5	0.89	0.82	0.89	0.78	0.77	1.05	0.87	0.80	0.91
	20	0.93	0.86	0.93	0.79	0.76	1.15	0.90	0.80	0.94
VAR(20), Lasso	1	1.00	0.91	0.97	0.85	0.83	1.20	0.97	0.85	1.02
Post Lasso OLS	5	0.91	0.84	0.91	0.79	0.77	1.10	0.89	0.80	0.94
	20	0.84	0.81	0.82	0.79	0.76	1.07	0.85	0.79	0.87
VAR(20), adaptive Lasso	1	1.09	1.04	1.06	0.98	0.97	1.10	1.05	1.00	1.07
Initial estimator: Lasso	5	0.93	0.87	0.93	0.83	0.82	1.05	0.91	0.84	0.94
	20	1.00	0.95	1.00	0.81	0.79	1.21	0.97	0.87	1.01
VAR(20), Lasso	1	1.09	1.00	1.06	0.90	0.90	0.95	1.00	0.93	1.03
Log-matrix transform	5	0.87	0.84	0.87	0.76	0.76	0.81	0.83	0.79	0.86
	20	0.79	0.86	0.79	0.74	0.74	0.78	0.80	0.80	0.79
VAR(20), Lasso	1	1.06	1.00	1.06	0.90	0.90	1.12	1.05	0.93	1.05
Including SP 500	5	0.89	0.82	0.89	0.78	0.77	1.05	0.90	0.80	0.91
	20	0.93	0.86	0.93	0.78	0.75	1.14	0.93	0.80	0.95

TABLE 2. Forecasts precision measures with daily observations. Absolute measures (gray background) and relative to benchmark (white background). A: All equations, D: Diagonal equations, O: Off-diagonal equations.



Model	VAR(1)			VAR(20)		No-Change	DCC	EWMA
	Lasso	adaLasso	Lasso	Lasso	Lasso	Censored		
Estimator:	Init: Lasso		(Log Mat)	(Log Mat)				
Statistic								
Max weight	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
Min weight	-0.20	-0.20	-0.20	-0.20	-0.20	-0.20	-0.20	-0.20
Proportion of leverage	0.22	0.22	0.27	0.22	0.28	0.24	0.24	0.23
Average turnover	0.02	0.03	0.02	0.02	0.01	0.03	0.01	0.01
Average % return ( $\times 10^{-2}$ )	2.58	3.58	5.89	2.85	6.27	1.99	0.40	0.14
Accumulated % return	10.07	15.16	27.77	11.45	30.02	7.13	-0.48	-1.60
Standard deviation ( $\times 10^{-2}$ )	0.97	0.98	1.00	0.97	0.99	0.97	1.00	0.99
Sharpe ratio ( $\times 10^{-2}$ )	2.66	3.65	5.87	2.95	6.29	2.04	0.40	0.14
Diversification ratio	1.46	1.47	1.43	1.46	1.43	1.48	1.43	1.43
Economic Value $\gamma = 1$								
No-Change (censored)	1.50	4.08	10.25	2.21	11.32	–	–	–
DCC	5.73	8.41	14.84	6.46	15.95	–	–	–
EWMA	6.40	9.09	15.56	7.13	16.68	–	–	–
Economic Value $\gamma = 5$								
No-Change (censored)	1.54	4.00	9.92	2.27	11.06	–	–	–
DCC	6.13	8.71	14.89	6.90	16.08	–	–	–
EWMA	6.68	9.27	15.48	7.45	16.68	–	–	–
Economic Value $\gamma = 10$								
No-Change (censored)	1.58	3.91	9.50	2.35	10.74	–	–	–
DCC	6.64	9.08	14.95	7.45	16.25	–	–	–
EWMA	7.04	9.49	15.38	7.85	16.69	–	–	–

TABLE 3. Statistics for the daily portfolios.

FIGURE 3.  $\ell_2$  norm of the 1 step ahead forecast error (left panel) and average penalty parameter (right panel) selected by BIC. A separate measure is included for the variance equations (diagonal terms) and the covariance equations (off-diagonal terms).