ESTIMATION AND FORECASTING OF LARGE REALIZED COVARIANCE MATRICES AND PORTFOLIO CHOICE

LAURENT A. F. CALLOT

VU University Amsterdam, The Netherlands, CREATES, and the Tinbergen Institute.

ANDERS B. KOCK

Aarhus University, CREATES, Denmark.

MARCELO C. MEDEIROS

Department of Economics, Pontifical Catholic University of Rio de Janeiro, Brazil.

ABSTRACT. In this paper we consider modeling and forecasting of large realized covariance matrices by penalized vector autoregressive models. We propose using Lasso-type estimators to reduce the dimensionality to a manageable one and provide strong theoretical performance guarantees on the forecast capability of our procedure. To be precise, we show that we can forecast future 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 for the realized covariance matrices of the 30 Dow Jones stocks and find that these dynamics are not stable as the data is aggregated from the daily to the weekly and monthly frequency.

The theoretical performance guarantees on our forecasts are illustrated on the Dow Jones index. In particular, we can beat our benchmark by a wide margin at the longer forecast horizons. 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 get access to our forecasts.

KEYWORDS: Realized covariance; vector autoregression; shrinkage; Lasso; forecasting; portfolio allocation.

JEL CODES: C22

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 based on ultra-high-frequency data as, for example, the composite realized kernel of Lunde et al. (2013). 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, Kirby and Ostdiek(2001,2003). Furthermore, our procedure is able to unveil the dynamics governing the evolution of the covariance matrices and we use this property to investigate which sectors are instrumental in driving these dynamics. This is extremely important since we can answer questions such as: are certain sectors independent of other sectors? Are the variances of stocks mainly described by their own past or are there spill over effects across stocks and sectors? Are the covariances governed by the same dynamics as the variances? Finally, 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. 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 (2012). 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, and as stressed already, our method can also 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 to assess the economic value of our forecasts.

Note that we do not consider the estimation of the covariance matrix in high dimensions. There are several papers recently proposed in the literature to tackle this problem as, for example, Bickel and Levina (2008), Levina et al. (2008), Fan et al. (2008), Wang and Zou (2010), Bai and Shi (2011), Fan et al. (2011), Fan et al. (2012), Fan et al. (2012a), Hautsch et al. (2009), Fan et al. (2013), and Lunde et al. (2013) among many others. 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 for researchers, especially in high dimensions where the potential number of parameters to be estimated becomes intractable by standard time series techniques.

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 heterogenous autoregressive (HAR) model put forward by Corsi (2009). 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. Another example is Chiriac and Voev (2011) who consider a multivariate ARFIMA model to forecast realized measures of daily covariance matrices. As in Bauer and Vorkink (2011), only small covariance matrices can be modelled. Golosnoy et al. (2012) propose a Conditional Autoregressive Wishart (CAW) model for the analysis of realized covariance matrices which is a dynamic generalization of the work put forward by Gourieroux et al. (2009). As before, in their application the authors only consider five different assets.

Using high-frequency data for portfolio selection has been considered by many authors. The work of Fleming et al. (2003), which builds on Fleming et al. (2001), measures 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. (2011) 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 than previously believed. Although, these authors consider the same problem as we do here, their modelling approach is quite different.

The rest of the paper is organized as follows. Section 2 describes the problem setup, defines notation, and briefly presents the Lasso and some key assumptions. In Section 3 we present some theoretical results. The data set and computational issues are discussed in Section 4, variable selection and forecasting results are presented in Section 5, portfolio selection using the forecasted covariance matrices is discussed in section 6. Finally, Section 7 concludes the paper. All proofs are deferred to the 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 which will be used throughout the paper.

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 $\|x\|_{\ell_\infty} = \max_{1 \le i \le m} |x_i|$ denote ℓ_2, ℓ_1 and ℓ_∞ norms, respectively. When regarding the $m \times m$ matrix A as a linear operator from $\mathbb{R}^m$ to $\mathbb{R}^m$ equipped with either the ℓ_1- or the ℓ_2-norm, $\|A\|$ and $\|A\|_{\ell_1}$ denote the induced operator norms. $\|A\|_\infty$ shall 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 Σ_t denote the $n_T \times n_T$ population conditional covariance matrix as of time t when conditioning on the σ-field $\sigma(\{\Sigma_s : s < t\})$. Note that the dimension n_T of Σ_t is indexed by the sample size T. This reflects the fact that n_T may be large compared to T and hence standard asymptotics, which take the dimension n_T as a fixed number, may not accurately reflect the actual performance in finite samples.

As is clear from the above, the entries of the conditional covariance matrix are potentially a complicated function of past entries of the conditional covariance matrix. If we allowed conditioning on the infinite past of the conditional covariance process, each entry of Σ_t would be a function of infinitely many variables. Instead, we assume that Σ_t only depends on the p_T most recent values of Σ_t, namely $\Sigma_{t-1}, ..., \Sigma_{t-p_T}$. As each of these matrices has $(n_T + 1)n_T/2$ unique entries we still have that every entry of Σ_t can be a function of $n_T \frac{n_T+1}{2} p_T$ variables. 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 Σ_t, every entry of Σ_t can be a function of up to 4,650 variables. This is the case for all the 465 unique entries of Σ_t [1]. To make this manageable, we shall assume that every entry of Σ_t is a *linear* function of it past, so that Σ_t follows a VAR(p_T) process. This allows for rich dynamics as every entry of Σ_t can depend on several thousand past entries, leading to equations with potentially many more parameters than observations.

---

[1] Actually, we shall work with moels including up to 20 lags, thus having 9,300 variables per equation.

Formally, define $y_t = \operatorname{vech} \Sigma_t$. We assume that y_t follows a vector autoregression of order p_T, i.e.,

$$(1) \qquad y_t = \sum_{i=1}^{p_T} \Phi_i y_{t-i} + \epsilon_t, \ t = 1, ..., T$$

where $\Phi_i, \ i = 1, ..., p_T$ are the $k_T \times k_T$ dimensional parameter matrices with $k_T = n_T(n_T + 1)/2$. Henceforth we suppress the dependence of n_T, k_T and p_T on T to simplify notation.

Note that the number of parameters per equation, k, increases quadratically in n, as does the number of equations. So even for conditional covariance matrices Σ_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 the number of observations. 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 the sample size.

Even though each entry of Σ_t can in principle 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 Σ_t. Furthermore, it is likely that the intrasectoral dynamics are stronger than the intersectoral ones. For these reasons the $\Phi_i, i = 1, ..., p$ in 1 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, as we shall see, the Lasso can be used to unveil and disentangle the potentially complex dynamics of the sequence of realized covariance matrices.

It is convenient to write the model (1) in stacked form. To do so let $Z_t = (y'_{t-1}, ..., y'_{t-p})'$ be the $kp \times 1$ vector of explanatory variables at time t in each equation and $X = (Z_T, ..., Z_1)'$ the $T \times kp$ matrix of covariates for each equation. Let $y_i = (y_{T,i}, ..., y_{1,i})'$ be the $T \times 1$ vector of observations on the ith variable $(i = 1, ..., k)$ and $\epsilon_i = (\epsilon_{T,i}, ..., \epsilon_{1,i})'$ the corresponding vector of error terms. Finally, β_i^* is the kp dimensional parameter vector of true parameters for equation i which also implicitly depends on T. Hence, we may write (1) equivalently as

$$(2) \qquad y_i = X\beta_i^* + \epsilon_i, \ i = 1, ..., k$$

such that each equation in (1) may be modeled separately. In our application modeling the DJIA, every equation has $465 \times p$ explanatory variables. However it is likely that many of these variables are irrelevant in explaining the dynamics of y_i, leading to many entries of β_i^* being equal to zero. For each $i = 1, ..., k$ we denote by $s_i = |\{j : \beta_{i,j}^* \neq 0\}|$ the number of non-zero entries in β_i^*.

2.3. **The Lasso.** We next describe briefly how the Lasso can be used to estimate a system of large equations as outlined above. The theoretical properties of the Lasso have been studied extensively, see e.g. Zhao and Yu (2006), Meinshausen and Bühlmann (2006), Bickel et al. (2012), Bühlmann and Geer (2011), and Kock and Callot (2012) 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, we shall see that it can still serve as an effective screening device in these situations. Put differently, it can remove many irrelevant covariates while keeping the relevant ones and estimating the coefficients of these with high precision. We investigate the properties of the Lasso when applied to each equation $i = 1, ..., k$ separately. We also provide finite sample bounds on the systemwise forecasting performance. The Lasso estimates β_i^* in (2) by minimizing the following objective function

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

with respect to β_i where λ_T is a sequence to be defined exactly below. (3) is basically the least squares objective function plus an extra term penalizing parameters that are different from zero. In practice λ_T is often chosen either by cross validation or using an information criterion. An attractive feature of the Lasso is that it performs estimation and variable selection simultaneously. This is useful in high-dimensional models as ours in particular, 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 inevitable have to resort to *ad hoc* procedures. Finally, the Lasso is a convex minimization problem, resulting in fast estimation and variable selection. This is a necessity for our high-dimensional problem where 30*31/2=465 equations with several thousand variables in each equation have to be estimated for every forecast we make.

2.4. **Forecasting with the Lasso.** Once the parameter estimates $\hat{\beta}_i$ have been obtained for equation i, forecasting is simple and is done as usual in VAR models. The one-step ahead forecast is given by

$$\hat{y}_{i,T+1} = \hat{\beta}_i' Z_T.$$

Doing this for all $i = 1, ..., k$ results in a forecast $\hat{\Sigma}_{T+1}$ of the matrix Σ_{T+1}. Forecasts at longer horizons are obtained by iterating forward the estimated VAR model, replacing unknown future values of $y_{i,t}$ by their forecasts.

3. Theoretical Results

In this section we provide an upper bound on the maximal forecast error $\|\hat{\Sigma}_{T+1} - \Sigma_{T+1}\|_\infty$. As we will see next 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. To be precise, 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

$$\hat{\sigma}^2_{T+1} = w'\hat{\Sigma}_{T+1}w.$$

How close are $\hat{\sigma}^2_{T+1}$ and σ^2_{T+1} 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 $\|\hat{\Sigma}_{T+1} - \Sigma_{T+1}\|_\infty$. Thus, we need a precision guarantee for the forecasts of the Lasso. In particular, 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. (2012b), 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,*

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

Theorem 1 reveals that in the presence of a restriction on the exposure of the portfolio, an upper bound on $\|\hat{\Sigma}_{T+1} - \Sigma_{T+1}\|_\infty$ implies an upper bound on the distance between $\hat{\sigma}^2_{T+1}$ and σ^2_{T+1}. 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. It is also worth noting that the bound in Theorem 1 is generic in the sense that the forecast $\hat{\Sigma}_{T+1}$ does not need to come from a specific method. Any matrix $\hat{\Sigma}_{T+1}$ can be plugged in.

We shall now give an upper bound on $\|\hat{\Sigma}_{T+1} - \Sigma_{T+1}\|_\infty$ based on our VAR approach. To this end, let $\sigma^2_{i,y}$ denote the variance of $y_{t,i}$ and $\sigma^2_{i,\epsilon}$ the variance of $\epsilon_{t,i}$, $1 \leq i \leq k$. Then define $\eta_T = \max_{1 \leq i \leq k}(\sigma_{i,y} \vee \sigma_{i,\epsilon})$. Also, let $\bar{s} = \max(s_1, ..., s_k)$ denote the maximal number of non-zero coefficients in any equation of the vector autoregression.

Theorem 2. *Let $\lambda_T = \sqrt{8\ln(1+T)^5 \ln(1+k)^4 \ln(1+p)^2 \ln(k^2 p)\eta^4_T/T}$ and $0 < q < 1$. Under regularity conditions made precise in the appendix, one has*

$$(4) \qquad \left\|\hat{\Sigma}_{T+1} - \Sigma_{T+1}\right\|_\infty \leq \sqrt{2\eta^2_T \ln(k(p+1))\ln(T)} \left(\frac{16}{q\kappa^2}\bar{s}\lambda_T + 1\right)$$

with high probability (the exact probability, as well as the definition of $\kappa > 0$, are given in the appendix).

Theorem 2 gives an upper bound on the forecast error of $\hat{\Sigma}_{T+1}$ 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, the larger this upper bound since there are more elements which have to be forecasted. 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 Φ_i, $i = 1, ..., p$ were known, one would have the following upper bound on the forecast error:

Lemma 1. *Assume that Φ_i, $i = 1, ..., p$ are known. Then, under regularity conditions made precise in the appendix,*

$$\left\|\hat{\Sigma}_{T+1} - \Sigma_{T+1}\right\|_\infty \le \sqrt{2\sigma_{T,\epsilon}^2 \ln(k) \ln(T)}$$

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

Compared to the bound in Theorem 2 the bound in Lemma 1 has roughly removed the term $\sqrt{2\eta_T^2 \ln(k(p+1)) \ln(T)} \frac{16}{q\kappa^2} \bar{s}\lambda_T$ 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 unremovable source of forecast error. From an asymptotic point of view one will usually requires that $\bar{s}\lambda_T \to 0$ 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 on the forecast error of $\hat{\sigma}_T^2$.

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

$$\left|\hat{\sigma}_{T+1}^2 - \sigma_{T+1}^2\right| \le \sqrt{2\eta_T^2 \ln(k(p+1)) \ln(T)} \left(\frac{16}{q\kappa^2}\bar{s}\lambda_T + 1\right)(1+c)^2$$

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

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 $c \to \infty$, the upper bound in the above display tends to ∞ 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 $\{w \in \mathbb{R}^n : \|w\|_{\ell_1} \le 1 + c\}$, 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*

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

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

As $\hat{w} = \arg\min_{w \in \mathbb{R}^n : \|w\|_{\ell_1} \le 1+c, \sum_{i=1}^n w_i=1} w'\hat{\Sigma}_{T+1}w$ 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 $\hat{w}$.

4. DATA AND IMPLEMENTATION

In this section we discuss the data used in the empirical application as well as details of the implementation of the estimation and forecasting strategies.

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[2]. The daily realized covariances are constructed from 5 minutes returns. The stocks can be classified in 8 broad categories highlighted in Table 1.

Basic Materials	Technology	Consumer Cyclical	Consumer Non-cyclical
2	4	3	7
Energy	Financial	Industrial	Communication
2	3	5	4

TABLE 1. Number of stocks per category. 30 Dow Jones stocks.

Prior to any sort of estimation or forecasting, 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 nearest five preceding and following non-flagged matrices. Using this cleaning scheme, the flagged matrices are concentrated in October 2008. The flash crashes of the 6th of May 2010 and the 9th 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 for horizons greater than 1, and all forecast errors are computed based on the de-transformed forecasts. We consider 3 levels of aggregation of the data: daily, weekly, and monthly. The 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. The weekly models are estimated using a rolling window of 263 observations resulting in 27 forecasts. The first weekly forecast is made for the first week of 2011. All monthly models are estimated using a rolling window of 60 observations resulting in 7 forecasts of which the first is made January 2011. We forecast with Vector autoregressive (VAR) and random walk (No-Change) models. The estimators are either

---

[2]We are grateful to Asger Lunde for kindly providing us with the data.

the Lasso, the adaptive Lasso (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 are carried out using `R` and the `lassovar` package based on the `glmnet` algorithm which is an implementation of the coordinate descent algorithm of Friedman et al. (2010). All the results are reproducible using the supplementary material provided.

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 λ is

$$BIC_i(\lambda) = T \times \log\left(\hat{\epsilon}'_{\lambda,i}\hat{\epsilon}_{\lambda,i}\right) + \sum_{j=1}^{kp} \mathbb{1}\left(\hat{\beta}^\lambda_{ij} \neq 0\right)\log(T).$$

where $\hat{\epsilon}_{\lambda,i}$ is the estimated vector of error terms corresponding to a penalty of λ. The results of the Lasso are compared to the ones of the adaptive Lasso. Letting $J(\hat{\beta}_i) = \{j \in \mathbb{R}^{kp} : \hat{\beta}_{i,j} \neq 0\}$ denote the indices of the coefficients in the i'th equation deemed zero by the Lasso, the adaptive Lasso estimates β_i^* by minimizing the following objective function

$$(5) \qquad \tilde{L}(\beta_i) = \frac{1}{T}\left\|y_i - X_{J(\hat{\beta}_i)}\beta_{i,J(\hat{\beta}_i)}\right\|^2 + 2\lambda_T \sum_{j \in J(\hat{\beta}_i)} \frac{|\beta_{i,j}|}{|\hat{\beta}_{i,j}|}, \ i = 1, ..., k$$

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 $\hat{\beta}_{i,j}$ is likely to be small by equation 9 (in the appendix) and consistency of the Lasso. Hence, $1/\hat{\beta}_{i,j}$ is large and the penalty on $\beta_{i,j}$ is large. Conversely, if $\beta_{i,j}^* \neq 0$, $\hat{\beta}_{i,j}$ is not too close to zero and the penalty is small.

We also consider a post Lasso least squares estimation (Belloni and Chernozhukov (2013)), that is, estimating the model by OLS after using the Lasso for variable selection. This post Lasso OLS has the benefit of not being subject to the bias introduced in the Lasso estimates due to penalization. Kock and Callot (2012) shows that the post Lasso estimator reduces the estimation and forecast error in a VAR framework.

5. DRIVING DYNAMICS AND FORECASTS
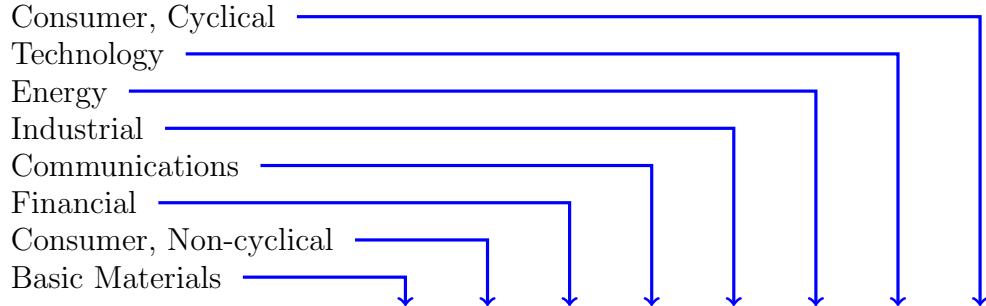
This section reports our empirical findings. We begin by studying the dynamic behavior of the sequence of realized covariance matrices by investigating which variables drive the dynamics of the realized covariance matrices. In particular, we investigate whether intrasectoral dependencies are stronger than intersectoral ones, and whether some stocks are important in explaining the volatility of a large fraction of the stocks in the Dow Jones index. We also investigate how the dynamics of the realized covariance matrices differ across levels of aggregation of the data (daily, weekly, monthly). We address these questions next.

5.1. **Driving dynamics of the realized covariance matrix.** The results in this section are based on a VAR(1) estimated by the Lasso. Thus, each of the 465 equations to be estimated has 465 variables (plus a constant). Table 2 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. Thus, the numbers are always between zero and one. 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. While it is easy to assign variances and covariances between two variables from the same category to that category, 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 inter-sectoral covariance terms enter as explanatory variables in two rows in Table 2. For that reason, the off-diagonal explanatory variables have to be interpreted with a lot of care and we will mainly focus on the diagonal terms.

The selected models in Table 2 for the variance terms (diagonal equations) are quite sparse. When explaining the dynamics of the variance equations, lags of variance terms are much more important than lags of covariance terms. Actually, the most important dynamics for 7 out of 8 sectors are the intra sectoral ones. 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 at the daily level of aggregation.

The bottom panel of Table 2 investigates the driving forces of the covariance terms. As mentioned already, these should be interpreted with a bit more care than the diagonal forces and we shall thus be brief in our comments. For 6 out of eight sectors it is lagged variances of that sector which are chosen most often in order to explain the dynamics of the corresponding sector. However, the off-diagonal terms are now chosen more often than above. Furthermore, the fraction of times the off-diagonal terms are chosen is naturally lower than the corresponding number for the diagonal terms as there are many more off-diagonal terms.

To complement the selection frequencies in Table 2 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. Consider first the VAR(1). The dimensions of the diagonal equations are remarkably stable, changing by only a couple of units throughout the sample. Furthermore, the diagonal equations are quite sparse only including between 11 and 13 variables out of a possible 465. The off-diagonal equations are less sparse and

Consumer, Cyclical
Technology
Energy
Industrial
Communications
Financial
Consumer, Non-cyclical
Basic Materials

	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 2. Fraction of variables selected from each sector (in row) when modeling variables from the sector in the columns. The numbers are averages across all the equations belonging to the sector in the row and averages across the number of estimations. The model is a VAR(1) estimated by the Lasso at the daily level of aggregation. The upper panel considers the dynamics of the variances while the bottom panel considers the dynamics of the covariances.
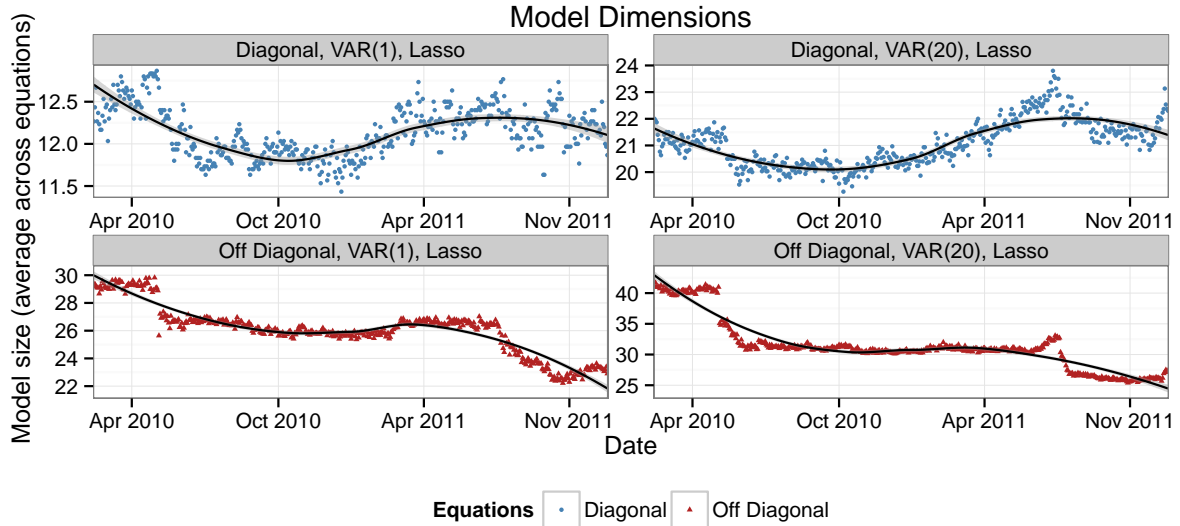
FIGURE 1. Average Equation size of the VAR(1) and VAR(20) estimated by the Lasso on censored data with 90% confidence intervals. A separate count is included for the variances (diagonal terms) and the covariances (off-diagonal terms).

their dimension, while being stable most of the time, drop around two of the periods of high volatility in May 2010 and August 2011. To be precise, the first sharp decrease in the model size at the beginning of the plot is at May 6, 2010 and is due to the flash crash on that day resulting in a clear structural break in the model. The second decrease happens around August 2011 and takes place in another period of extreme volatility. It is of interest that the variance equations are less affected by these two extreme periods.

The dimensions of the VAR(20) (right panels) are at most twice as large as those of the VAR(1) despite the twenty-fold increase in the number of potential covariates. The development of the model sizes for the variances and the covariances are very similar to the ones found for the VAR(1). The drop in the size of the equations for the covariances actually becomes even clearer.

In the left panels of Figure 2 we investigate how the ℓ_2 forecast error of the one-step ahead forecasts behaves throughout the sample, and in particular how it reacts to the periods of extreme volatility. As can be expected from the development of the model sizes over time described above the forecast errors of the covariances react much stronger to the periods of extreme volatility than the forecast errors of the variances. The right panels of Figure 2 explain the decrease in the model size of the covariance equations from a modelling point of view: the decrease in the model size is associated with an increase in the penalty parameter λ which determines the amount of shrinkage. This finding is valid for the VAR(1) as well as the VAR(20).

We also investigated the model sizes when the VAR(1) was estimated on the uncensored data set. Interestingly, the variance equations were hardly affected by this. On the other
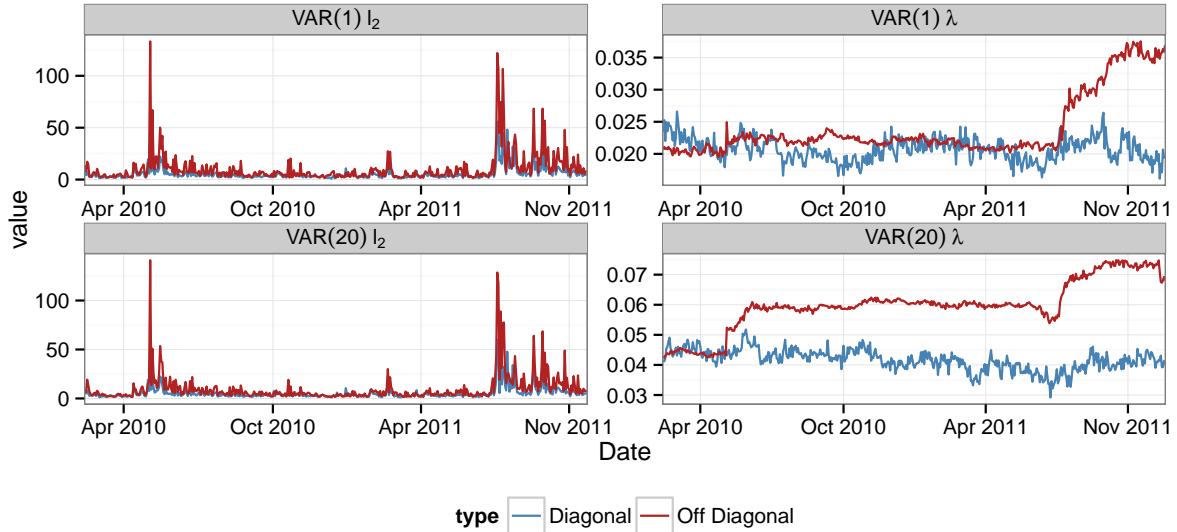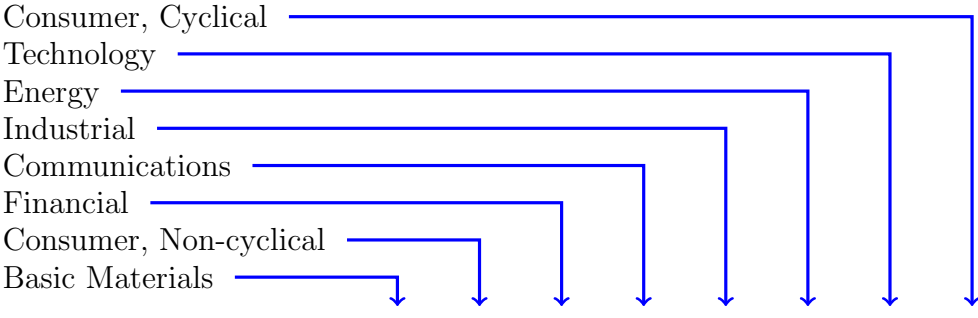
FIGURE 2. ℓ_2 norm of the 1 step ahead forecast error (left panel) and average penalty parameter (right panel) selected by BIC. A separate measure is included for the variance equations (diagonal terms) and the covariance equations (off-diagonal terms).

hand, the model sizes were now around 130 for the covariance equations with an even sharper drop of around 30 around the flash crash on May 6, 2010.

Turning to the weekly horizon, the upper panel of Table 3 reveals that the variances are still mainly explained by lags of their own past even though the pattern is slightly less pronounced. Furthermore, as opposed to the daily level of aggregation, no sector includes all of its own past. The off-diagonal terms still do not play much of a role in explaining the dynamics of the diagonal terms. As is revealed from the bottom panel of Table 3, a higher fraction of diagonal than off-diagonal variables is chosen when modeling covariances. This is in accordance with Table 2. At this frequency the variances of from the Energy and Financial sectors are frequently retained to explain the volatility in other sectors.

Table 4 reports the selection frequencies at the monthly level of aggregation. A remarkable fact is that the diagonal pattern for the explanatory variables of the variances has now disappeared almost entirely. 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 sum up, as the level of aggregation of the data is increased, the diagonal pattern becomes less and less clear. At the daily level of aggregation the variances were to a high degree explained by their own past such that the intrasectoral dynamics were dominant. At the monthly horizon the intersectoral dynamics are more pronounced and the volatility in the sample is driven by the financial and energy sectors. The weekly level of aggregation strikes a middle ground between these two extremes.

The top diagram maps each sector (rows on the left) to the corresponding column of the tables below, in the column order: Basic Materials, Consumer Non-cyclical, Financial, Communications, Industrial, Energy, Technology, Consumer Cyclical.

	Basic Materials	Consumer, Non-cyclical	Financial	Communications	Industrial	Energy	Technology	Consumer, Cyclical
Variance Equations								
Lagged variance								
Basic Materials	0.75	0.13	0.08	0.35	0.08	0.00	0.41	0.19
Consumer, Non-cyclical	0.12	0.43	0.08	0.15	0.04	0.00	0.02	0.08
Financial	0.17	0.54	0.78	0.33	0.66	0.19	0.18	0.51
Communications	0.26	0.08	0.04	0.27	0.04	0.00	0.08	0.04
Industrial	0.00	0.00	0.00	0.00	0.69	0.00	0.01	0.00
Energy	0.47	0.20	0.19	0.20	0.00	0.81	0.31	0.32
Technology	0.22	0.17	0.01	0.07	0.00	0.02	0.55	0.00
Consumer, Cyclical	0.19	0.32	0.06	0.20	0.25	0.06	0.23	0.64
Lagged covariance								
Basic Materials	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Consumer, Non-cyclical	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Financial	0.00	0.01	0.02	0.01	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	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
Energy	0.01	0.01	0.00	0.01	0.00	0.00	0.01	0.01
Technology	0.00	0.00	0.00	0.00	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	0.43	0.12	0.09	0.11	0.11	0.27	0.13	0.07
Consumer, Non-cyclical	0.07	0.26	0.07	0.10	0.03	0.17	0.08	0.08
Financial	0.12	0.33	0.79	0.20	0.06	0.46	0.16	0.23
Communications	0.23	0.17	0.10	0.28	0.17	0.14	0.24	0.18
Industrial	0.04	0.01	0.00	0.02	0.02	0.00	0.03	0.00
Energy	0.25	0.14	0.31	0.04	0.02	0.38	0.18	0.17
Technology	0.29	0.14	0.06	0.12	0.10	0.07	0.36	0.16
Consumer, Cyclical	0.24	0.26	0.18	0.15	0.37	0.29	0.18	0.51
Lagged covariance								
Basic Materials	0.06	0.05	0.03	0.07	0.05	0.06	0.05	0.04
Consumer, Non-cyclical	0.03	0.05	0.04	0.04	0.04	0.04	0.04	0.04
Financial	0.06	0.07	0.10	0.07	0.04	0.05	0.05	0.06
Communications	0.04	0.04	0.05	0.06	0.04	0.05	0.05	0.04
Industrial	0.06	0.05	0.05	0.07	0.14	0.04	0.05	0.05
Energy	0.04	0.05	0.03	0.05	0.04	0.10	0.07	0.05
Technology	0.03	0.02	0.03	0.03	0.02	0.03	0.05	0.02
Consumer, Cyclical	0.03	0.03	0.03	0.03	0.02	0.03	0.04	0.04

TABLE 3. Fraction of variables selection from each sector (in rows) when modeling variables from the sector in the columns. The numbers are averages across all the equations belonging to the sector in the row and averages across the number of estimations. The model is a VAR(1) estimated by the Lasso at the weekly level of aggregation. The upper panel considers the dynamics of the variances while the bottom panel considers the dynamics of the covariances.

Variance Equations

	Sector								
Lagged variance	Basic Materials	0.10	0.19	0.21	0.30	0.48	0.06	0.27	0.34
	Consumer, Non-cyclical	0.00	0.10	0.01	0.01	0.06	0.00	0.00	0.02
	Financial	0.00	0.52	0.65	0.40	0.68	0.10	0.10	0.35
	Communications	0.14	0.14	0.09	0.19	0.19	0.08	0.17	0.15
	Industrial	0.00	0.03	0.01	0.01	0.00	0.04	0.16	0.05
	Energy	0.57	0.35	0.28	0.31	0.10	0.61	0.41	0.37
	Technology	0.12	0.06	0.00	0.04	0.03	0.01	0.23	0.01
	Consumer, Cyclical	0.01	0.06	0.09	0.13	0.08	0.02	0.11	0.09
Lagged covariance	Basic Materials	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Consumer, Non-cyclical	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Financial	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00
	Communications	0.00	0.01	0.00	0.01	0.01	0.00	0.01	0.01
	Industrial	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Energy	0.01	0.01	0.00	0.01	0.00	0.01	0.01	0.01
	Technology	0.00	0.01	0.00	0.01	0.00	0.00	0.01	0.01
	Consumer, Cyclical	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Covariance Equations

	Sector								
Lagged variance	Basic Materials	0.69	0.89	0.94	0.78	0.75	0.90	0.75	0.77
	Consumer, Non-cyclical	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.00
	Financial	0.18	0.35	0.49	0.26	0.09	0.10	0.14	0.24
	Communications	0.29	0.40	0.42	0.34	0.29	0.41	0.35	0.39
	Industrial	0.39	0.13	0.11	0.05	0.05	0.38	0.26	0.15
	Energy	0.62	0.63	0.61	0.52	0.17	0.70	0.63	0.57
	Technology	0.05	0.01	0.02	0.01	0.02	0.02	0.05	0.00
	Consumer, Cyclical	0.42	0.45	0.50	0.42	0.38	0.44	0.40	0.37
Lagged covariance	Basic Materials	0.01	0.01	0.02	0.01	0.00	0.03	0.01	0.01
	Consumer, Non-cyclical	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Financial	0.00	0.01	0.01	0.01	0.00	0.01	0.00	0.00
	Communications	0.05	0.05	0.04	0.05	0.05	0.05	0.05	0.05
	Industrial	0.02	0.03	0.05	0.03	0.01	0.07	0.03	0.03
	Energy	0.02	0.02	0.02	0.02	0.03	0.05	0.02	0.02
	Technology	0.04	0.06	0.06	0.06	0.03	0.05	0.04	0.04
	Consumer, Cyclical	0.02	0.03	0.04	0.02	0.02	0.04	0.02	0.02

TABLE 4. Fraction of variables selection from each sector (in rows) when modeling variables from the sector in the columns. The numbers are averages across all the equations belonging to the sector in the row and averages across the number of estimations. The model is a VAR(1) estimated by the Lasso at the monthly level of aggregation. The upper panel considers the dynamics of the variances while the bottom panel considers the dynamics of the covariances.

Finally, it can be observed from the bottom panel of Table 4 that the basic materials sector is completely dominant in explaining the dynamics of the covariance terms irrespective of the which sector these are related to. This is rather surprising in light of the fact that the basic materials sector did not play any particular role for the variance terms.
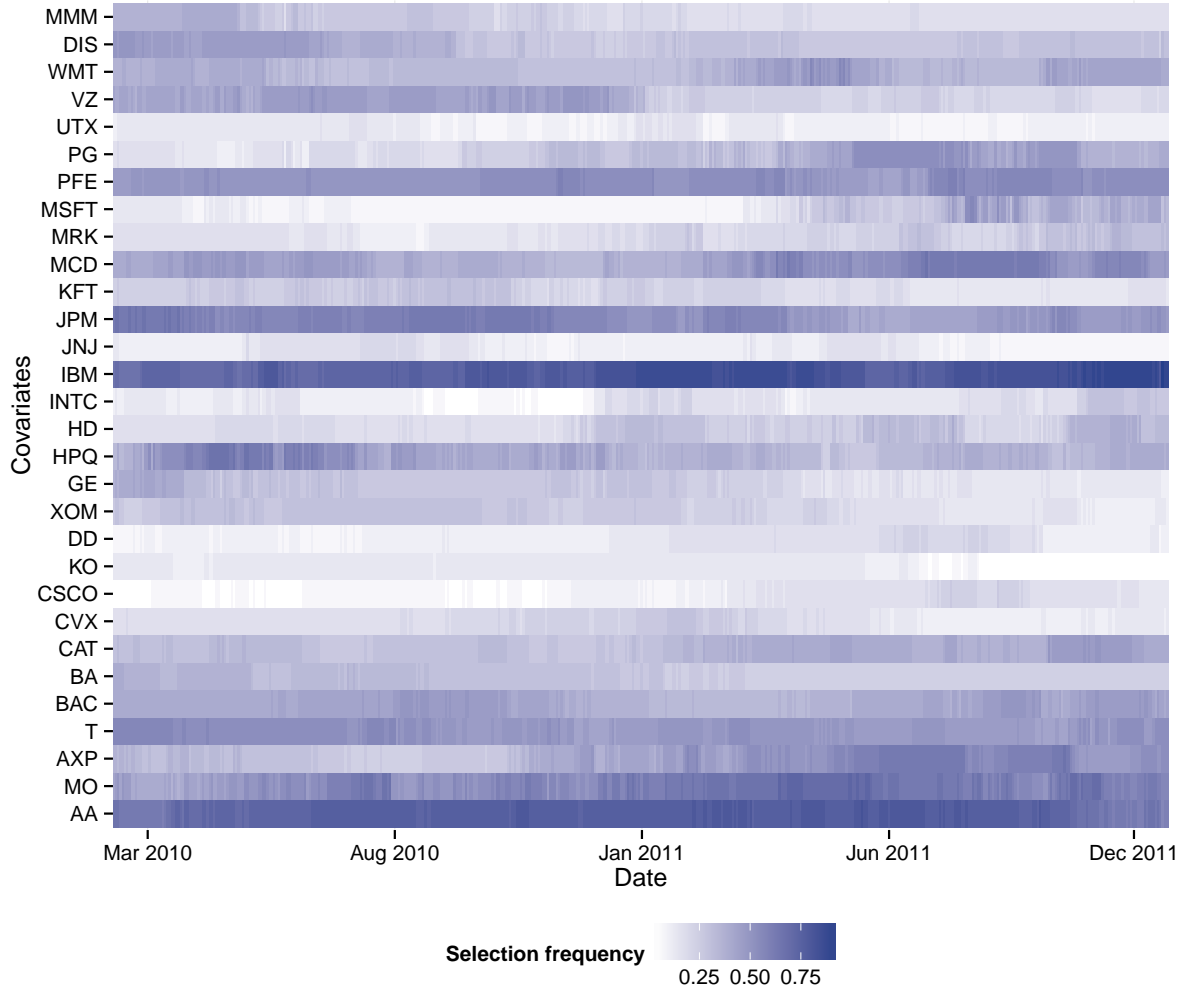

FIGURE 3. Selection frequency of the stocks in the columns. At each point in time on the x-axis we calculate the selection frequence across the 30 variance equations of each of the stocks on the y-axis. The results are based on a VAR(1) model estimated by the Lasso at the daily level of aggregation.

Figure 3 investigates the dynamics of the thirty variance equations in more detail at the daily level of aggregation. To be precise, it indicates in what fraction of the 30 equations the lagged variance of the the stock indicated on the y-axis is chosen as an explanatory variable at each point in time on the x-axis. The first thing that can be noticed is the relative stability of the selected variables over the 455 forecasts: the selection frequency of each variable varies very little over time.

From Figure 3 it is also seen that the variance of IBM is the one which is most often chosen by the Lasso for explaining the dynamics of the realized variances. Actually, IBM is chosen in almost 80 percent of the equations on average. On the other hand,

UTX (United Technologies Corporation) is the stock whose variance is least important in explaining the realized variances. More precisely, lagged variances of IBM are chosen between seven and eight times as often as those of UTX.
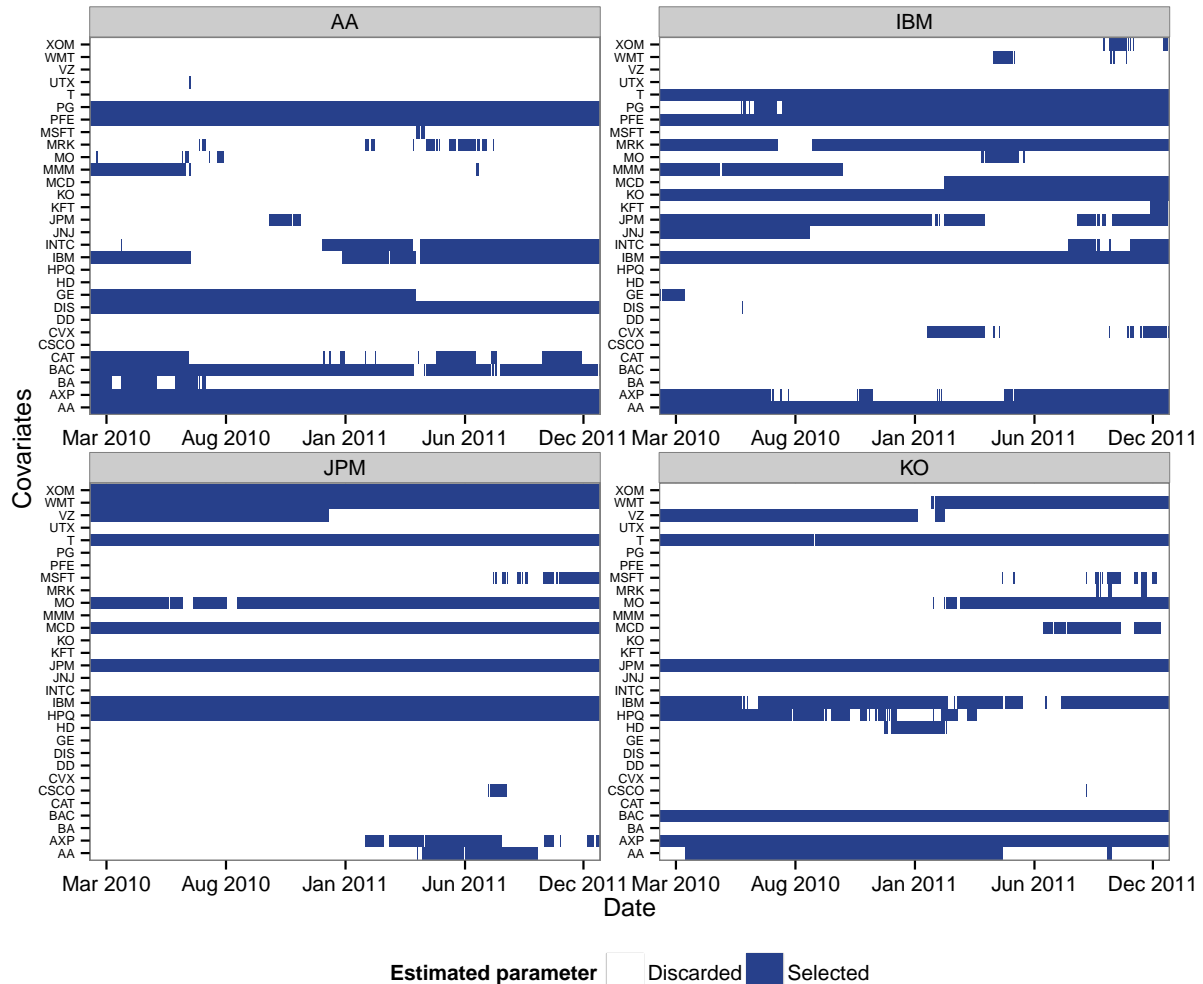
FIGURE 4. For the variance for each of the four stocks Alcoa (AA), IBM, JP Morgan (JPM), and Coca Cola (KO), the figure contains the variances (indicated on the y-axis) selected into the equation describing their dynamics at each point in time on the x-axis. The results are based on a VAR(1) model estimated by the Lasso at the daily level of aggregation.

To shed even further light on the findings in Table 2 we also consider the variable selection pattern for 4 individual stocks in more detail. Figure 4 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.

Overall, 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[3]. All stocks except Coca-Cola always include their own lagged variance – a fact which is in line with the findings in Table 2. Furthermore, with a few exceptions, this pattern also emerges when investigating the corresponding plots for the 26 remaining Dow Jones stocks.

5.2. **Forecasts.** In this section we forecast the realized covariance matrices using various VAR models. We are particularly interested in how the forecast precision varies over different degrees of aggregation of the data. We consider the daily, weekly, and monthly level of aggregation and estimate the VAR model by the Lasso, the adaptive Lasso as well as the post-Lasso. Furthermore, we investigate the effect of allowing for greater time series dependence in the covariance matrices by including more lags in the model.

Let $\hat{e}_{T+h}$ denote $vech\left(\hat{\Sigma}_{T+h} - \Sigma_{T+h}\right)$ where $\hat{\Sigma}_{T+h}$ denotes the forecast of Σ_{T+h} as of time T at horizon $h \geq 1$. Thus, $\hat{e}_{T+h}$ is the vector of forecast error 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 ℓ_2-forecast error (ℓ_2): $\frac{1}{T_n-T_1+1}\sum_{T=T_1}^{T_n}\|\hat{e}_{T+h}\|$. We shall employ this measure for the whole matrix, the variances, or the covariance terms. 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): $\frac{1}{T_n-T_1+1}\sum_{T=T_1}^{T_n} median\left(\hat{e}_{T+h}\right)$. This measure is considered for the whole matrix, the variances, and the covariances. It is more robust to outliers than the ℓ_2-norm.

(3) Average maximal absolute forecast error (AMaxAFE): $\frac{1}{T_n-T_1+1}\sum_{T=T_1}^{T_n} max\left(\hat{e}_{T+h}\right)$. 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 forecasted covariance matrices produced by our models 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 in Section 6. To overcome the ill-conditioning and ensure that the forecasted covariance matrices are positive definite, we regularize the forecasted covariance matrices by eigenvalue cleaning. We apply eigenvalue cleaning to every matrix that has eigenvalues strictly smaller than 0 or a condition number greater than $10n_T$ as in Hautsch et al. (2009).

Write the spectral decomposition of the forecasted covariance matrix $\widehat{\Sigma}_t = \widehat{V}_t'\widehat{\Lambda}_t\widehat{V}_t$ where $\widehat{V}_t$ is the matrix of eigenvectors and $\widehat{\Lambda}_t$ a diagonal matrix with the N eigenvalues $\widehat{\lambda}_{it}$ on its diagonal. Let $\lambda_{\mathsf{mp}t} = \min\{\widehat{\lambda}_{it}|\widehat{\lambda}_{it} > 0\}$, replace all the $\widehat{\lambda}_{it} < \lambda_{\mathsf{mp}t}$ by $\lambda_{\mathsf{mp}t}$ and define the diagonal matrix $\widetilde{\Lambda}_t$ with the cleaned eigenvalues on its diagonal. The regularized forecast matrix $\widetilde{\Sigma}_t = \widehat{V}_t'\widetilde{\Lambda}_t\widehat{V}_t$ is by construction positive definite.

---

[3]Here it should be kept in mind that we only focus on which lagged variances are chosen for explaining the the variances. Had we also investigated how often covariance terms were chosen, the fraction of variables included would have been even lower as can also concluded from Table 2.

| Model | h | AMedAFE | | | AMaxAFE | | | ℓ_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
	20	0.58	0.98	0.57	5.12	5.12	2.22	18.05	9.25	15.17
No-Change	1	0.36	0.65	0.35	6.1	6.1	2.34	14.91	9.37	10.83
Uncensored	5	0.55	0.96	0.53	7.23	7.23	2.87	19.89	11.74	15.19
	20	0.66	1.16	0.65	7.85	7.85	3.18	22.93	13.1	17.94
VAR(1), Lasso	1	0.37	0.61	0.37	3.34	3.32	1.72	11.98	5.93	10.21
	5	0.44	0.73	0.43	3.77	3.64	2.25	14.25	6.82	12.27
	20	0.69	0.96	0.68	4.37	4.03	3.16	19.98	8.11	18.07
VAR(1), Lasso	1	0.34	0.55	0.33	3.08	3.04	1.76	11.26	5.4	9.72
Post Lasso OLS	5	0.45	0.73	0.44	3.8	3.68	2.23	14.39	6.87	12.36
	20	0.61	0.93	0.6	4.34	4.09	2.94	18.55	8.06	16.43
VAR(1), adaptive Lasso	1	0.37	0.62	0.37	3.46	3.44	1.81	12.21	6.07	10.4
Initial estimator: Lasso	5	0.44	0.74	0.44	3.88	3.78	2.32	14.49	6.93	12.52
	20	0.62	0.98	0.61	4.45	4.18	3.13	19.44	8.38	17.3
VAR(1), Lasso	1	0.38	0.61	0.37	3.34	3.32	1.74	12.04	5.93	10.27
Not regularized	5	0.45	0.73	0.44	3.83	3.63	2.32	14.35	6.81	12.38
	20	0.69	0.95	0.68	4.54	4	3.34	20.01	8.04	18.12
VAR(20), Lasso	1	0.35	0.57	0.35	3.19	3.16	1.62	11.35	5.59	9.66
	5	0.41	0.65	0.4	3.54	3.46	2.01	13.09	6.28	11.25
	20	0.54	0.84	0.53	4.03	3.87	2.56	16.29	7.44	14.3
VAR(20), Lasso	1	0.33	0.52	0.32	3.01	2.92	1.76	10.88	5.09	9.44
Post Lasso OLS	5	0.42	0.66	0.41	3.56	3.48	2.1	13.43	6.31	11.65
	20	0.49	0.79	0.47	4.02	3.9	2.38	15.29	7.27	13.25
VAR(20), adaptive Lasso	1	0.36	0.59	0.35	3.45	3.44	1.61	11.76	5.98	9.89
Initial estimator: Lasso	5	0.43	0.69	0.42	3.75	3.72	2.01	13.62	6.66	11.65
	20	0.58	0.93	0.57	4.16	4.04	2.68	17.49	8.03	15.33

TABLE 5. The table contains the measures of forecast precision discussed in the main text at the daily level of aggregation. The forecast horizons are h=1, 5 and 20 days ahead. A: Measure applied on all elements of the matrix. D: Measure applied on all diagonal (variances) elements of the matrix. O: Measure applied on all off-diagonal (covariances) elements of the matrix.

All the forecast errors analyzed in the tables below are computed from the regularized forecasts. For comparison purpose we also report results on unregularized forecasts from a VAR(1) estimated by Lasso, highlighted in red in the tables.

Table 5 contains the forecast results for the daily level of aggregation. We will mainly focus on the extent to which it is possible to beat the no-change (random walk) forecasts as these can be challenging benchmarks in particular at short horizons. In order to reassess the importance of filtering the data prior to forecasting we have also included no-change forecasts for the censored as well as the uncensored series. As can be seen, the forecast errors are uniformly larger for the uncensored data irrespective of which measure

of precision one uses. This is of course not surprising as the uncensored data contains some big spikes which entail big changes in the data and thus lower the precision of the no-change forecasts.

It is quiet interesting that no matter which model or estimation method is used for forecasting, the variances seem to be harder to forecast than the covariances. This can be witnessed from the larger median and maximal average forecast error of the forecasts of the variances. Keeping in mind that there are only 30 variances to be forecasted, as opposed to 435 covariances, this is also confirmed by the fact that the ℓ_2-forecast error is not much lower for the variances than for the covariances. We believe this is an important and new finding.

Furthermore, it is encouraging that it does not make a big difference whether one estimates the VAR models by the Lasso, post-Lasso or the adaptive Lasso. The post-Lasso might be slightly more precise at the daily level of aggregation and thus we shall focus mainly on that for now.

At all forecast horizons, the post-Lasso has a lower median forecast error for the variances than the no-change forecasts. On the other hand, it is less precise than these when it comes to forecasting the covariances. The overall median forecast error should be interpreted with care as there are many more covariance terms than variances. Therefore, the overall median mainly reflects the median forecast error of the off-diagonal terms of the covariance matrix, as can also be witnessed by the relative closeness of these two numbers in Table 5. 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 25 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.

Figure 1 showed that the models for the covariance terms were in general much larger than those for the variances in a VAR(1). This indicates that the the Lasso tries to compensate for some unexplained dynamics by including many variables. In order to better capture the dynamics of the off-diagonal terms we also experimented with gradually including more lags. Thus, Table 5 also contains the results for VAR(20) models. As was the case for the VAR(1) models, the post-Lasso again seems to be the dominant estimation procedure. However, and more importantly, it now provides more precise forecasts of the covariance terms than the no-change forecasts. Furthermore, this does not come at the price of worse forecasts for the variances. In fact, these forecasts also become more precise. The median forecast error is now as good as the one for the no-change forecasts overall, for both the diagonal terms as well as the off-diagonal terms. On the other hand, the maximal forecast error for the covariances, while improved, still remains slightly above the one for the no-change forecasts. The ℓ_2-forecast errors are lower than the ones from the no-change forecasts.

Table 6 contains the forecast errors at the weekly level of aggregation. It confirms the finding for the daily forecast that the variances are harder to forecast than the covariances.

Model	h	AMedAFE			AMaxAFE			ℓ_2		
		A	D	O	A	D	O	A	D	O
No-Change	1	0.11	0.29	0.11	1.62	1.62	0.53	4.49	2.97	3.2
	10	0.59	1	0.58	4.54	4.54	1.86	16.83	8.35	14.36
	26	0.96	1.46	0.94	7.26	7.26	3.26	27.44	13.05	24
VAR(1), Lasso	1	0.16	0.31	0.15	1.58	1.58	0.81	5.49	3.01	4.49
	10	0.62	0.96	0.6	4.5	4.32	2.36	17.43	7.97	15.4
	26	0.69	1.15	0.68	6.47	6.37	2.79	21.45	10.85	18.23
VAR(1), Lasso Post Lasso OLS	1	0.15	0.26	0.14	1.46	1.41	1.01	5.25	2.69	4.43
	10	0.69	1.17	0.68	5.11	4.86	2.81	21.01	9.28	18.38
	26	Inf	Inf	Inf	Inf	Inf	Inf	Inf	Inf	Inf
VAR(1), adaptive Lasso Initial estimator: Lasso	1	0.15	0.32	0.15	1.66	1.66	0.78	5.54	3.13	4.46
	10	0.65	0.96	0.64	4.65	4.3	2.58	18.51	7.99	16.51
	26	0.7	1.14	0.69	6.47	6.3	2.91	21.6	10.79	18.41
VAR(1), Lasso Not regularized	1	0.16	0.31	0.15	1.59	1.59	0.82	5.51	3.02	4.51
	10	0.62	0.96	0.6	4.5	4.32	2.37	17.44	7.98	15.4
	26	0.69	1.16	0.68	6.73	6.55	2.9	21.59	11.04	18.26
VAR(5), Lasso	1	0.15	0.3	0.14	1.52	1.52	0.82	5.25	2.93	4.26
	10	0.62	0.95	0.61	4.37	4.34	2.05	17.2	7.9	15.16
	26	0.7	1.15	0.68	6.56	6.48	2.69	21.57	10.96	18.27
VAR(5), Lasso Post Lasso OLS	1	0.13	0.26	0.12	1.44	1.39	0.98	4.97	2.64	4.17
	10	0.59	0.91	0.58	4.52	4.41	2.49	17.09	7.88	15.09
	26	0.78	1.27	0.76	7.4	7.29	3.3	24.42	12.34	20.86
VAR(5), adaptive Lasso Initial estimator: Lasso	1	0.15	0.32	0.14	1.66	1.65	0.83	5.39	3.11	4.29
	10	0.64	0.96	0.63	4.42	4.23	2.33	17.97	7.87	16
	26	0.7	1.15	0.69	6.54	6.41	2.79	21.66	10.93	18.38

TABLE 6. The table contains the measures of forecast precision discussed in the main text at the weekly level of aggregation. The forecast horizons are h=1, 10 and 26 weeks ahead. A: Measure applied on all elements of the matrix. D: Measure applied on all diagonal (variances) elements of the matrix. O: Measure applied on all off-diagonal (covariances) elements of the matrix.

Furthermore, this seems to be the case whether we consider median or maximal absolute forecasts errors.

Turning to the VAR(1) models, the plain Lasso seems to yield the most precise forecasts. It has a slightly lower median forecast error for the variances than the no-change forecasts and is much more precise when forecasting the covariances 26 periods ahead. This is also seen from the ℓ_2-forecast errors which are around 25 percent lower for the Lasso than for the no-change forecasts.

In contrast to the daily level of aggregation, there no longer seems to be any benefit from adding lags. Neither the forecasts of the variances nor the forecasts of the covariances improve for any of the estimation procedures when including five lags instead of one.

Finally, Table 7 confirms that the variances are harder to forecast than the covariances even at the monthly level of aggregation. Of the three VAR(1) models, the adaptive Lasso

| Model | h | AMedAFE | | | AMAaxFE | | | ℓ_2 | | |
		A	D	O	A	D	O	A	D	O
No-Change	1	0.11	0.22	0.1	1.28	1.28	0.67	4.34	2.56	3.38
	2	0.44	0.71	0.43	2.99	2.99	1.33	12.16	5.8	10.6
	6	0.87	1.29	0.86	6.15	6.15	2.84	24.49	11.34	21.64
VAR(1), Lasso	1	0.18	0.23	0.18	1.22	0.99	1.05	5.53	2.16	5.03
	2	0.65	0.78	0.65	3.5	3.26	2.35	17.38	6.14	16.18
	6	0.81	1.12	0.8	5.91	5.58	2.92	24.03	9.95	21.26
VAR(1), Lasso Post Lasso OLS	1	0.29	0.33	0.29	1.82	1.36	1.74	9.03	2.99	8.45
	2	0.7	0.87	0.69	3.87	3.61	2.59	19.35	6.92	17.93
	6	1.14	2.17	1.13	8.09	7.54	6.57	39.24	16.92	34.51
VAR(1), adaptive Lasso Initial estimator: Lasso	1	0.15	0.23	0.15	1.13	1.06	0.89	5	2.2	4.45
	2	0.53	0.79	0.52	3.44	3.25	1.88	14.61	6.16	13.22
	6	0.74	1.1	0.73	5.57	5.5	2.53	21.53	9.82	18.9
VAR(1), Lasso Not regularized	1	0.29	0.29	0.29	2.31	1.53	2.27	9.57	2.98	9.02
	2	0.86	0.7	0.87	4.84	3.25	3.5	23.76	5.87	22.63
	6	2.31	1.35	2.36	13.04	6.44	11.33	65.96	11.71	63.68
VAR(5), Lasso	1	0.14	0.25	0.14	1.05	0.99	0.89	4.75	2.12	4.24
	2	0.44	0.73	0.43	3.26	3.26	1.5	12.41	5.87	10.88
	6	0.73	1.12	0.71	6.09	6.09	2.66	21.31	10.28	18.48
VAR(5), Lasso Post Lasso OLS	1	0.15	0.26	0.15	1.33	1.2	1.17	5.87	2.49	5.28
	2	0.52	0.73	0.51	3.07	2.91	1.86	14.44	5.72	13.2
	6	1.08	1.3	1.07	6.02	5.93	4.06	29.61	10.96	27.33
VAR(5), adaptive Lasso Initial estimator: Lasso	1	0.13	0.23	0.13	1.15	1.04	0.95	4.71	2.1	4.18
	2	0.45	0.73	0.44	3.32	3.32	1.55	12.69	5.93	11.17
	6	0.67	1.08	0.66	5.93	5.93	2.32	19.9	10.06	16.9

TABLE 7. The table contains the measures of forecast precision discussed in the main text at the monthly level of aggregation. The forecast horizons are h=1, 2 and 6 months ahead. A: Measure applied on all elements of the matrix. D: Measure applied on all diagonal (variances) elements of the matrix. O: Measure applied on all off-diagonal (covariances) elements of the matrix.

seems to deliver the most precise forecasts though the performances of all three models are reasonably similar. At the monthly level of aggregation, it seems quite difficult to beat the no-change forecasts with the 6-month horizon being an exception. Adding more lags improves the precision of the forecasts slightly and the 6-month forecast are now much more precise for the adaptive Lasso than for the no-change forecasts.

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 smaller performance gains obtained from including lags at the weekly and monthly level of aggregation can partly be attributed to the fact that there are much fewer observations at these levels of aggregation such that including more parameters becomes much more costly. The particular version of the Lasso used to estimate the VAR model does not seem overly important, which makes the procedure rather robust.

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 the framework developed in Fleming, Kirby, and Ostdiek (2001,2003) (FKO, hereafter). 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. To be precise, the investor utility, at each point in time, is given by

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

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

As in the previous sections we consider three different frequencies of observations: daily, weekly and monthly. The portfolios are rebalanced in each time period as follows. Let r_{t+1}, $\widehat{\mu}_{t+1}$, and $\widehat{\Sigma}_{t+1}$ denote, respectively, an $n \times 1$ vector of observed 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. $\widehat{\Sigma}_{t+1}$ is computed by the methods described earlier while $\widehat{\mu}_{t+1}$ is computed by a moving average of N time periods, where $N = 100$ for daily data, $N = 15$ for weekly data, and $N = 4$ for monthly data.

The investor's problem at $t = t_0, \ldots, T - 1$ 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:

$$\widehat{\omega}_{t+1} = \arg\min_{\omega_{t+1}} \omega'_{t+1}\widehat{\Sigma}_{t+1}\omega_{t+1}$$

$$\text{s.t.} \quad \omega'_{t+1}\widehat{\mu}_{t+1} = \mu_{\text{target}}$$

$$(7) \qquad \sum_{i=1}^{n} \omega_{it+1} = 1$$

$$\sum_{i=1}^{n} |\omega_{it+1}|\mathsf{I}(\omega_{it} < 0) \leq 0.30$$

$$|\omega_{it+1}| \leq 0.20,$$

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

The optimal weights, $\widehat{\omega}_{t+1}$, which solve the optimization problem described in (7) can be viewed as a function of the target expected return μ_{target}, the conditional expected return $\widehat{\mu}_{t+1}$ and the conditional covariance matrix forecast $\widehat{\Sigma}_{t+1}$. We impose two additional restrictions on the optimal weights. First, we allow the maximum leverage to be 30%

(corresponding to $c = 0.6$ in Section 3). This yields much stabler portfolios. However, the optimization problem does not have a closed form solution anymore and a numerical constrained minimization routine must be used. 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. Thus, the one-step ahead forecasts seem most relevant in this context.

Results. Tables 8–10 present several summary statistics. Overall, eight different forecasting models have been considered: six Lasso-based VARs and a benchmark based on no-change forecasts. The tables display the following statistics.

(1) Average weight: $\frac{1}{n(T-t_0)} \sum_{t=t_0+1}^{T} \sum_{i=1}^{n} \widehat{\omega}_{it}$. This is the average weight across all assets and all time periods.

(2) Max weight: $\max_{t_0+1 \leq t \leq T} \max_{1 \leq i \leq n} (\widehat{\omega}_{it})$ for $t = t_0+1, \ldots, T$ and $= 1, \ldots, N$. This is the maximum weight over all assets and all time periods.

(3) Min weight: $\min_{t_0+1 \leq t \leq T} \min_{1 \leq i \leq n} (\widehat{\omega}_{it})$ for $t = t_0 + 1, \ldots, T$ and $= 1, \ldots, N$. This is the minimum weight over all assets and all time periods.

(4) Average leverage: $\frac{1}{n(T-t_0)} \sum_{t=t_0+1}^{T} \sum_{i=1}^{n} |\widehat{\omega}_{it}| \mathsf{I}(\widehat{\omega}_{it} < 0)$. This is the average of negative weights (short positions) across all assets and all time periods.

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

(6) Average turnover: $\frac{1}{n(T-t_0)} \sum_{t=t_0+1}^{T} \sum_{i=1}^{n} |\widehat{\omega}_{it} - \widehat{\omega}_{it}^{\mathsf{hold}}|$, where $\widehat{\omega}_{it}^{\mathsf{hold}} = \widehat{\omega}_{it-1} \frac{(1+r_{it-1})}{1+r_{pt-1}}$. The turnover measures the average change in the portfolio composition (portfolio weights). $\omega_{it}^{\mathsf{hold}}$, $i = 1, \ldots, n$, 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.

(7) Average return: $\mu_p = \frac{1}{(T-t_0)} \sum_{t=t_0+1}^{T} r_{pt} = \frac{1}{(T-t_0)} \sum_{t=t_0+1}^{T} \widehat{\omega}_t' r_t$. Note this is the out-of-sample average portfolio return as at each period t the weights are selected based only on information up to time $t-1$.

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

(9) 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.

(10) 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.

(11) Average diversification ratio: $\frac{1}{(T-t_0)} \sum_{t=t_0+1}^{T} \frac{\sum_{i=1}^{n} \widehat{\omega}_{it} \sigma_{it}}{\sigma_{pt}}$, where $\sigma_{pt} = \widehat{\omega}_t \Sigma_t \widehat{\omega}_t'$. This ratio is the portfolio's weighted average the individual asset volatility (σ_{it}) to its actual total volatility (σ_{pt}). In a well-diversified portfolio the total risk of such

portfolio is often less than the weighted-average risk of its individual parts and diversification arises when returns are not perfectly correlated.

(12) Economic value: the economic value is the value of Δ such that, for different portfolios p_1 and p_2, we have

$$\sum_{t=t_0+1}^{T} U(r_{p_1 t}) = \sum_{t=t_0+1}^{T} U(r_{p_2 t} - \Delta).$$

The economic value represents the maximum return the investor would be willing to sacrifice each time period in order to capture 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 Δ as an annualized basis point fee, for three levels of relative risk aversion ($\gamma = 1, 5, 10$).

Table 8 contains the results at the daily frequency and several conclusions emerge. First, the average return over the forecasting period is virtually the same for all models considered. On the other hand, the accumulated return over the 445 trading days is much higher than the one for the benchmark for four out of six Lasso based models. More specifically, the VAR(1) plain Lasso is the model which delivers the highest accumulated returns, 31.46%. It is followed by the VAR(20) estimated with adaptive Lasso, 28.21%. The benchmark alternative has an accumulated return between 23.9%. These are big differences, especially when considering that the annual target return (approximately 252 trading days) is 10%. In terms of standard deviation, all models have similar figures. One possible reason for this is the imposition of the leverage and maximum weight constraints, which restrict the exposition of the portfolio. The VAR(1) plain Lasso model is also the

Model	VAR(1)			VAR(20)			No-Change
Estimator: Statistic	Lasso	Post Lasso OLS	adaLasso Init: Lasso	Lasso	Post Lasso OLS	adaLasso Init: Lasso	Censored
Average weight	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Max weight	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
Average leverage	0.27	0.26	0.27	0.28	0.27	0.27	0.27
Proportion of leverage	0.26	0.26	0.26	0.26	0.26	0.25	0.28
Average turnover	0.02	0.03	0.02	0.02	0.02	0.02	0.03
Average return ($\times 10^{-4}$)	7.23	6.39	5.96	5.50	5.67	6.69	5.88
Accumulated return	31.5	26.7	24.1	21.5	22.5	28.2	23.9
Standard deviation	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Sharpe ratio	0.05	0.04	0.04	0.04	0.04	0.04	0.04
Diversification ratio	1.50	1.50	1.50	1.50	1.50	1.50	1.51
Economic Value $\gamma = 1$ No-Change (censored)	3.36	1.27	0.10	-1.04	-0.59	1.94	–
Economic Value $\gamma = 5$ No-Change (censored)	2.90	1.10	-0.36	-1.38	-0.81	1.42	–
Economic Value $\gamma = 10$ No-Change (censored)	2.31	0.90	-0.92	-1.81	-1.08	0.77	–

TABLE 8. Statistics for the daily portfolios.

Model	VAR(1)			VAR(5)			No-Change
Estimator:	Lasso	Post Lasso	adaLasso	Lasso	Post Lasso	adaLasso	
Statistic		OLS	Init: Lasso		OLS	Init: Lasso	Uncensored
Average weight	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Max weight	0.19	0.20	0.19	0.19	0.20	0.18	0.19
Min weight	-0.17	-0.20	-0.18	-0.19	-0.20	-0.20	-0.20
Average leverage	0.28	0.28	0.29	0.28	0.29	0.29	0.28
Proportion of leverage	0.22	0.24	0.21	0.20	0.23	0.21	0.23
Average turnover	0.02	0.02	0.02	0.02	0.03	0.02	0.02
Average return ($\times 10^{-3}$)	2.01	2.57	2.60	2.51	2.44	3.22	2.75
Accumulated return	9.69	13.0	13.1	12.6	12.2	16.9	14.1
Standard deviation	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Sharpe ratio	0.09	0.12	0.12	0.12	0.12	0.16	0.13
Diversification ratio	1.44	1.44	1.42	1.42	1.44	1.41	1.50
Economic Value $\gamma = 1$							
No-Change	-17.5	-4.89	-4.33	-6.26	-7.83	12.5	–
Economic Value $\gamma = 5$							
No-Change	-19.3	-6.32	-6.03	-6.79	-8.46	13.0	–
Economic Value $\gamma = 10$							
No-Change	-21.6	-8.13	-8.16	-7.46	-9.25	13.7	–

TABLE 9. Statistics for the weekly portfolios.

one with the highest Sharpe ratio. In terms of diversification, all models perform in a very similar way. In addition, all portfolios seem to have the same leverage. However, the turnover of some the Lasso based models is lower than the corresponding turnover for the benchmark portfolios which can be useful in the presence of large transaction costs and explains some of the return differences. Finally, and more importantly, we also report the annualized economic value of the Lasso based portfolios with respect to the two no-change alternatives. The economic value is the amount, in terms, of annualized returns, an investor will be willing to pay to move from the benchmark portfolio to the Lasso based ones. As we can see from the table, the VAR(1) plain Lasso model delivers a positive annualized economic value of up to 3.6%. Furthermore, the higher the risk aversion of the investor, the lower the economic value of the Lasso based models. This is in line with the previous results as the differences in terms of the standard deviations of the portfolios are not big and most of the economic value comes from the differences in the accumulated returns over the out-of-sample period.

From Table 9 we can highlight the following findings. First, the average return over the forecasting period is virtually the same for all models considered. On the other hand, only the model estimated by adaptive Lasso has an accumulated return over the 52 trading weeks higher than the the benchmark (no-change) alternative. More specifically, the VAR(5) adaptive Lasso is the model which delivers the highest accumulated returns and the largest Sharpe ratio, 16.9% and 0.16, respectively. In terms of economic value, the VAR(5) adaptive Lasso is the only model to beat the benchmark. Finally, it is interesting to notice that reducing the frequency of rebalancing (from daily to weekly) reduces the accumulated returns in a very significant way.

Model	VAR(1)			VAR(5)			No-Change
Estimator: Statistic	Lasso	Post Lasso OLS	adaLasso Init: Lasso	Lasso	Post Lasso OLS	adaLasso Init: Lasso	
Average weight	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Max weight	0.20	0.20	0.20	0.19	0.19	0.18	0.16
Min weight	-0.18	-0.18	-0.18	-0.20	-0.19	-0.20	-0.14
Average leverage	0.29	0.28	0.29	0.30	0.29	0.30	0.11
Proportion of leverage	0.21	0.24	0.21	0.19	0.24	0.20	0.23
Average turnover	0.04	0.05	0.04	0.04	0.04	0.03	0.03
Average return ($\times 10^{-3}$)	6.55	3.87	6.44	8.38	10.7	10.7	8.57
Accumulated return	7.74	3.93	7.52	9.96	13.2	13.1	10.4
Standard deviation	0.03	0.04	0.03	0.03	0.03	0.03	0.03
Sharpe ratio	0.25	0.10	0.22	0.27	0.39	0.37	0.34
Diversification ratio	1.42	1.44	1.44	1.46	1.41	1.47	1.48
Economic Value $\gamma = 1$ No-Change	-40.4	-72.3	-42.9	-8.08	68.7	66.2	–
Economic Value $\gamma = 5$ No-Change	-41.8	-81.3	-47.6	-20.7	60.2	52.3	–
Economic Value $\gamma = 10$ No-Change	-43.7	-89.2	-53.5	-35.2	49.5	35.1	–

TABLE 10. Statistics for the monthly portfolios.

Table 10 displays the results concerning the monthly rebalancing. The overall picture is very similar to the weekly case.

7. CONCLUSIONS AND FURTHER WORK

In this paper 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 knowing the true model 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 intra-sectoral while at the longer horizons the dynamics were often described by the energy and financial sector. A forecast exercise revealed that our procedure outperforms our benchmark – a finding which is particularly pronounced at the long forecast horizons.

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 of aggregation our Lasso based forecast have an economic value of up to 3.6% per year.

8. APPENDIX

8.1. **Notation.** We shall use the following notation throughout the appendix. Let $J_i = \{j : \beta_{i,j}^* \neq 0\} \subseteq \{1, ..., kp\}$ denote the set of non-zero parameters in equation i and $s_i = |J_i|$ its cardinality. $\bar{s} = \max\{s_1, ..., s_k\}$.

For any vector δ in $\mathbb{R}^n$ and a subset $J \subseteq \{1, ..., n\}$ we shall let δ_J denote the vector consisting only of those elements of δ indexed by J.

Proof of Theorem 1. By the definition of σ_t^2 and $\hat{\sigma}_t^2$ one has

$$\left| \hat{\sigma}_t^2 - \sigma_t^2 \right| = \left| w'(\hat{\Sigma}_t - \Sigma_t) w \right| \le \left\| (\hat{\Sigma}_t - \Sigma_t) w \right\|_{\ell_\infty} \|w\|_{\ell_1}$$

$$\le \left\| \hat{\Sigma}_t - \Sigma_t \right\|_\infty \|w\|_{\ell_1}^2 \le \left\| \hat{\Sigma}_t - \Sigma_t \right\|_\infty (1 + c)^2$$

$\square$

Before we proceed to proving Theorem 2 we need to introduce the so-called restricted eigenvalue condition. To this end, let $\Gamma = E(Z_t Z_t')$ denote the population covariance matrix of Z_t. Note that Γ does not depend on t as we shall assume stationarity in Lemma 2 below. Then, we define

$$(8) \qquad \kappa_i^2 = \kappa_\Gamma^2(s_i) = \min \left\{ \frac{\delta' \Gamma \delta}{\|\delta_R\|^2} : |R| \le s_i, \ \delta \in \mathbb{R}^{kp} \setminus \{0\}, \ \|\delta_{R^c}\|_{\ell_1} \le 3\|\delta_R\|_{\ell_1} \right\}$$

In the sequel we shall assume that κ_i^2 is bounded away from zero for $i = 1, ..., k$. Note that this is the case in particular if Γ has full rank which is a rather innocent assumption as Γ is the *population* covariance matrix. For more details on the restricted eigenvalue condition we refer to Kock and Callot (2012).

Next, we recall the following result which is an extract of Theorem 2 in Kock and Callot (2012) and of vital importance in the proof of Theorem 2.

Lemma 2 (Theorem 2 in Kock and Callot (2012)). *Assume that $\kappa^2 := \min_{1 \le i \le k} \kappa_i^2$ is bounded away from zero and that $\epsilon_t \sim N_k(0, \Omega)$. All roots of $|I_k - \sum_{j=1}^p \Phi_j z^j|$ lie outside the unit circle. Let $\lambda_T = \sqrt{8 \ln(1+T)^5 \ln(1+k)^4 \ln(1+p)^2 \ln(k^2 p) \eta_T^4 / T}$ and $0 < q < 1$. Then with probability at least $1 - 2(k^2 p)^{1-\ln(1+T)} - 2(1+T)^{-1/A} - \pi_q(s_i)$ [4] the following inequalities hold for all $i = 1, ..., k$ for some positive constant A.*

$$(9) \qquad \left\| \hat{\beta}_i - \beta_i^* \right\|_{\ell_1} \le \frac{16}{q \kappa_i^2} s_i \lambda_T$$

Furthermore, all the above statements hold on one and the same set which has probability at least $1 - 2(k^2 p)^{1-\ln(1+T)} - 2(1+T)^{-1/A} - \pi_q(\bar{s})$.

Proof of Theorem 2. Since

$$\left\| \hat{\Sigma}_{T+1} - \Sigma_{T+1} \right\|_\infty = \left\| \text{vech} \, \hat{\Sigma}_{T+1} - \text{vech} \, \Sigma_{T+1} \right\|_\infty = \|\hat{y}_{T+1} - y_{T+1}\|_{\ell_\infty}$$

we shall bound each entry of $\hat{y}_{T+1} - y_{T+1}$. By assumption

$$y_{T+1,i} = Z_{T+1}' \beta_i^* + \epsilon_{T+1,i}$$

---

[4] $\pi_q(s) = 4k^2 p^2 \exp\left(\frac{-\zeta T}{s^2 \log(T)(\log(k^2 p^2)+1)} \right) + 2(k^2 p^2)^{1-\log(T)}$ for $\zeta = \frac{(1-q)^2 \kappa_i^4}{4 \cdot 16^3 (\|\Gamma\| \sum_{i=0}^T \|F^i\|)^2}$, F is the companion matrix of the VAR model. Under general condition, made clear in Kock and Callot (2012), $\pi_q(s)$ tends to zero.

while

$$\hat{y}_{T+1,i} = Z'_{T+1}\hat{\beta}_i$$

such that

$$\left| y_{T+1,i} - \hat{y}_{T+1,i} \right| = \left| Z'_{T+1}(\beta^*_i - \hat{\beta}_i) + \epsilon_{T+1,i} \right| \leq \|Z_{T+1}\|_{\ell_\infty} \|\hat{\beta}_i - \beta^*_i\|_{\ell_1} + \left| \epsilon_{T+1,i} \right|$$

Using Lemma 2 this yields that

$$\left| y_{T+1,i} - \hat{y}_{T+1,i} \right| \leq \|Z_{T+1}\|_{\ell_\infty} \frac{16}{q\kappa_i^2} s_i \lambda_T + \left| \epsilon_{T+1,i} \right| \text{ for all } i = 1, ..., k$$

with probability at least $1 - 2(k^2 p)^{1-\ln(1+T)} - 2(1+T)^{-1/A} - \pi_q(\bar{s})$. Next, note that by the gaussianity of the covariates and error terms $P(|y_{T-l,i}| \geq x) \leq 2e^{-x^2/2\eta_T^2}$ for all $1 \leq i \leq k$ and $1 \leq l \leq p$ and $P(|\epsilon_{T+1,i}| \geq x) \leq 2e^{-x^2/2\eta_T^2}$ for all $1 \leq i \leq k$. This implies[5]

$$P\left(\|Z_{T+1}\|_{\ell_\infty} \vee \max_{1\leq i \leq k} |\epsilon_{T+1,i}| \geq L \right) \leq 2kpe^{-L^2/2\eta_T^2} + 2ke^{-L^2/2\eta_T^2} = 2k(p+1)e^{-L^2/2\eta_T^2}$$

Choosing $L^2 = 2\eta_T^2 \ln(k(p+1))\ln(T)$ yields

(10) $$P\left(\|Z_{T+1}\|_{\ell_\infty} \vee \max_{1\leq i \leq k} |\epsilon_{T+1,i}| \geq L \right) \leq 2[k(p+1)]^{1-\ln(T)}$$

and so

$$\left| y_{T+1,i} - \hat{y}_{T+1,i} \right| \leq \sqrt{2\eta_T^2 \ln(k(p+1))\ln(T)} \left(\frac{16}{q\kappa_i^2} s_i \lambda_T + 1 \right)$$

$$\leq \sqrt{2\eta_T^2 \ln(k(p+1))\ln(T)} \left(\frac{16}{q\kappa^2} \bar{s} \lambda_T + 1 \right) \text{ for all } i = 1, \ldots, k$$

with probability at least $1 - 2(k^2 p)^{1-\ln(1+T)} - 2(1+T)^{-1/A} - \pi_q(\bar{s}) - 2[k(p+1)]^{1-\ln(T)}$. □

Proof of Lemma 1. When the Φ_i $i = 1, ..., p$ is known $\left\|\hat{\Sigma}_{T+1} - \Sigma_{T+1}\right\|_\infty = \|\epsilon_{T+1}\|_{\ell_\infty}$. Next, for all $1 \leq i \leq n$ one has[6] $P\left(|\epsilon_{T+1,i}| > L\right) \leq e^{-L^2/\sigma_{T,\epsilon}^2}$. Hence,

$$P\left(\max_{1\leq i \leq k} |\epsilon_{T+1,i}| > L \right) \leq ke^{-L^2/\sigma_{T,\epsilon}^2}$$

such that choosing $L^2 = \sigma_{T,\epsilon}^2 \log(k)\ln(T)$ yields

$$P\left(\max_{1\leq i \leq k} |\epsilon_{T+1,i}| \geq L \right) \leq k^{1-\ln(T)}$$

Therefore, with probability at least $1 - k^{1-\ln(T)}$ one has $\left\|\hat{\Sigma}_{T+1} - \Sigma_{T+1}\right\|_\infty \leq \sigma_{T,\epsilon}\sqrt{\ln(k)\ln(T)}$. □

Proof of Corollary 1. Combine Theorems 1 and 2. □

Proof of Corollary 2. Combine Theorems 1 and 2 noting that the former is valid uniformly over $\{w \in \mathbb{R}^n : \|w\|_{\ell_1} \leq 1 + c\}$. □

---

[5]The stationarity of y_t and ϵ_t implies that η_T^2 is also an upper bound on the variance of $\epsilon_{T+1,i}$ for all $i = 1, ..., k$.

[6]The stationarity of ϵ_t implies that $\sigma_{T,\epsilon}^2$ can also be used as an upper bound on the variance of all entries of ϵ_{T+1}.

References

Audrino, F. and S. Knaus (2014). Lassoing the har model: A model selection perspective on realized volatility dynamics. *Econometric Reviews*.

Bai, J. and S. Shi (2011). Estimating high dimensional covariance matrices and its applications. *Annals of Economics and Finance 12*, 199–215.

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 (2012). Simultaneous analysis of Lasso and Dantzig selector. *Journal of Applied Econometrics 27*, 525–703.

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

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.

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

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 (Methodological) 75*, 603–680.

Fan, J., J. Zhang, and K. Yu (2012a). Vast portfolio selection with gross-exposure constraints. *Journal of American Statistical Association 107*, 592–606.

Fan, J., J. Zhang, and K. Yu (2012b). Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association 107*(498), 592–606.

Fleming, J., C. Kirby, and B. Ostdiek (2001). The economic value of volatility timing. *Journal of Finance 56*, 329–352.

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.

Gourieroux, C., J. Jasiak, and R. Sufana (2009). The wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics 150*, 167–181.

Hautsch, N., L. Kyj, and P. Malec (2011). Do high-frequency data improve high-dimensional portfolio allocations? Discussion Paper 2011-059, SFB 649, Humboldt-Universität zu Berlin.

Hautsch, N., L. Kyj, and R. Oomen (2009). A blocking and regularization approach to high-dimensional realized covariance estimation. *The Annals of Statistics 37*, 625–645.

Kock, A. and L. Callot (2012). Oracle inequalities for high dimensional vector autoregressions. Research Paper 12, CREATES, Aarhus University.

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 (2013). Econometric analysis of vast covariance matrices using composite realized kernels. Discussion paper, University of Oxford.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance 7*, 77–91.

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.