Forecasting Large Realized Covariance Matrices: The Benefits of Factor Models and Shrinkage *

Diego S. de Brito
Department of Economics
PUC-Rio

Marcelo C. Medeiros
Department of Economics
PUC-Rio

Ruy M. Ribeiro
Department of Economics
PUC-Rio

First version: October 2017

This version: January 2018

Abstract

We propose a model to forecast very large realized covariance matrices of returns, applying it to the constituents of the S&P 500 on a daily basis. To deal with the curse of dimensionality, we decompose the return covariance matrix using standard firm-level factors (e.g. size, value, profitability) and use sectoral restrictions in the residual covariance matrix. This restricted model is then estimated using Vector Heterogeneous Autoregressive (VHAR) models estimated with the Least Absolute Shrinkage and Selection Operator (LASSO). Our methodology improves forecasting precision relative to standard benchmarks and leads to better estimates of the minimum variance portfolios.

Keywords: Realized covariance, factor models, shrinkage, Lasso, forecasting, portfolio allocation, big data.

*We thank Michael Wolf for insightful comments and for providing us with the Matlab code to estimate the NL-DCC and NL-BEKK models. We are also thankful to the seminar participants at the Third International Workshop In Financial Econometrics, Arraial d'Ajuda, Brazil, October 2017, for providing us with many comments and suggestions.

Brito: diegosiebra13@gmail.com. Medeiros: mcm@econ.puc-rio.br. Ribeiro: ruy.ribeiro@econ.puc-rio.br.

1 Introduction

The goal of this paper is to construct models based on factors and shrinkage methods to forecast large-dimensional and time-varying realized measures of daily covariance matrices of returns on financial assets. Realized measures of a covariance matrix are estimates, based on intraday returns, of the integrated covariance matrix of a multivariate diffusion process. One example of such estimator, used in this paper, is the composite realized kernel method recently introduced by Lunde, Shephard, and Sheppard 2016. Our proposed model is evaluated both in terms of its forecasting ability and, more importantly, in terms of several performance measures in a conditional mean-variance portfolio allocation problem.

Modeling and forecasting the covariance matrix of financial assets is essential for portfolio allocation and risk management. Moreover, it is an established empirical fact that (conditional) covariance matrices vary considerably over time. Natural ways to model such dynamics is to either use multivariate generalizations of the ARCH/GARCH family of models as proposed by Engle and Kroner (1995) or Engle (2002) or model directly a given realized measure of the covariance matrices by usual multivariate time-series models as in Bauer and Vornik (2011), Chiriac and Voev (2011) or Golosnoy, Gribisch, and Liesenfeld (2012). This is motivated by the close connection between the conditional covariance matrix and the integrated covariance of a multivariate diffusion process.

However, when the number of assets increases, the amount of parameters to be estimated gets very large. For instance, for a covariance matrix of n assets, there are $n(n+1)/2$ distinct entries to be modeled. If a vector autoregressive specification of order p, VAR(p), is used, then the total number of parameters will be $n(n+1)(p+1)/2$. Therefore, the curse of dimensionality precludes the application of the above referenced methods to moderately large covariance matrices and most of the previous studies in the literature focused on sets of less than ten assets.

More recently, based on the advances of modern statistical tools to handle high-dimensional models, new alternatives have been proposed in the literature in order to deal with a large number of assets. Callot, Kock, and Medeiros (2017) advocate the use of the Least Absolute Shrinkage and Selection Operator (LASSO) of Tibshirani (1996) and the adaptive LASSO of Zou (2006) to model the dynamics of the realized covariance of the constituents of the Dow Jones index by a large dimensional VAR. However, their modeling strategy is not able to handle sets of assets larger than the 30 used in the paper. Engle, Ledoit, and Wolf (2017) combine nonlinear shrinkage with the DCC-GARCH model and put forward a methodology where the dynamics huge latent conditional covariance matrices can be modeled.

In this paper we extend the results in Callot, Kock, and Medeiros (2017) by developing a modeling strategy which is able to handle high-dimensional sets of assets. More specifically, as Engle, Ledoit, and Wolf (2017), our method can be applied to hundreds or even thousands of assets with the difference that we will consider realized measures of the covariance model in opposition to latent ones as in Engle, Ledoit, and Wolf (2017). We show empirical evidence that considering realized measures of the covariance matrix drastically improve the forecasting ability and produce portfolios with better performance measures. Our model is based of a common factor structure to the vector of returns which induce a decomposition of the covariance matrix. Hence, the daily covariance matrix can be written into a covariance matrix for a low dimensional set of factors plus an idiosyncratic covariance matrix which is (almost) block-diagonal. The dynamics of the variances and covariances of the factors are modeled by a VHAR model estimated with LASSO. To guarantee positive definiteness of the forecasts we apply the log matrix transformation of Chiu, Leonard, and Tsui (1996). The daily factor loadings (betas) are computed with the high-frequency data. We show that these loading exhibit a high degree of long-range dependence and we model their dynamics by a HAR specification as well. Finally, the dynamics of each block of the idiosyncratic covariance matrix is model by a restricted autoregressive model estimated as well by LASSO. As far as we are concerned this is the first model which is able to describe the dynamics of huge realized covariance matrices. Fan, Furger, and Xiu (20016) also considered factors and realized covariances but their forecasting model is just a random walk and their results are more concerned with the estimation of the integrated covariance.

Despite the interesting challenges involved in the construction of realized measures (e.g. how to deal with microstructure noise), specially in the multivariate case (e.g. asynchronicity in transactions of different assets, which bias covariance estimations towards zero), this work will not follow this path. Instead, we focus solely on modeling and forecasting large realized covariance matrix of returns on hundred of financial assets which are estimated elsewhere.

The rest of paper is organized as follows. Section 2 describes the proposed model and the forecasting framework. In Section 4 we present the data and show a descriptive analysis. The forecasting results and the portfolio analysis are presented, respectively, in Sections 5 and 6. Finally, Section 7 concludes the paper. Supplementary results are shown in the Appendix.

2 Model

We base our methodology on the fact that realized covariance matrices are highly persistent over time, which suggests the use of an autoregressive model of a large order p, usually larger than 20. Defining $\boldsymbol{y}_t = \text{vech}(\boldsymbol{\Sigma}_t)$, where vech is the half-vectorization operator returning a vector with the unique entries of $\boldsymbol{\Sigma}_t$, one possible specification is:

$$\boldsymbol{y}_t = \boldsymbol{\omega} + \sum_{i=1}^{p} \boldsymbol{\Phi}_i \boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t, \tag{1}$$

where $\boldsymbol{\epsilon}_t$ is a zero-mean random noise.

While this model is sensible for a small number of assets, the number of parameters grows quadratically (curse of dimensionality). To circumvent this limitation, Callot, Kock, and Medeiros (2017) propose the use of penalized regressions (LASSO) as a way to deal with the large number of parameters. However, the direct use of penalized regressions is unfeasible for hundreds of assets. To illustrate, assume that the representation in (1) is used to model the covariance matrix for the constituents of the S&P 500 index[1], with 10 lags, that is $p = 10$. Since each matrix $\mathbf{\Sigma}_t$ has $n(n+1)/2$ unique entries, this configuration would result in $125,250$ equations with $10 \times 125,250$ variables each, plus constants. In this case, estimation is unfeasible even when using the LASSO.

In order to reduce dimensionality to a manageable one, we propose the use of a factor model as well as economic restrictions based on sector classifications and penalized regressions (Lasso).

2.1 Factor Model

Following the factor model discussed in Chamberlain and Rothschild (1983), the excess return on any asset i, $r_{i,t}$, satisfies:

$$r_{i,t}^e = \beta_{i1,t} f_{1,t} + \cdots + \beta_{iK,t} f_{K,t} + \varepsilon_{i,t} = \boldsymbol{\beta}_{i,t}' \boldsymbol{f}_t + \varepsilon_{i,t} \tag{2}$$

where $f_{1,t}, \cdots, f_{K,t}$ are the excess returns of K factors, $\beta_{ik,t}$, $k = 1, \ldots, K$, are factor loadings for asset i, and $\varepsilon_{i,t}$ is the error term. Notice that the factor loadings are variable over time. For N assets, the set of equations can be written in matrix form:

$$\boldsymbol{r}_t^e = \boldsymbol{B}_t' \boldsymbol{f}_t + \boldsymbol{\varepsilon}_t \tag{3}$$

where $\boldsymbol{B}_t$ is a $K \times N$ matrix of loadings, $\boldsymbol{r}_t$ is a $N \times 1$ vector of excess returns, and $\boldsymbol{\varepsilon}_t$ is a $N \times 1$ vector of idiosyncratic errors. Throughout we assume that $\mathbb{E}(\boldsymbol{\varepsilon}_t | \boldsymbol{f}_t) = \mathbf{0}$. The factors used in this work are linear combinations of returns constructed solely with the assets here considered, i.e. long-short stock portfolios where stocks that are part of our

[1] In our application, we restrict the analysis to stocks that remain in the index for the whole period of our sample. This reduces the number of stocks to 430. In terms of number of equations and potential predicting variables, our problem is still subject to the curse of dimensionality in the same order of magnitude as the illustration.

sample are sorted on firm characteristics. In matrix form, for all K factors:

$$\begin{pmatrix} f_{1,t} \\ \vdots \\ f_{K,t} \end{pmatrix} = \begin{pmatrix} w_{1,1} & \dots & w_{1,N} \\ \vdots & \ddots & \vdots \\ w_{K,1} & \dots & w_{K,N} \end{pmatrix} \begin{pmatrix} r_{1,t} \\ \vdots \\ r_{N,t} \end{pmatrix} \tag{4}$$

or

$$\boldsymbol{f}_t = \boldsymbol{W}'\boldsymbol{r}_t,$$

where weights are calculated based on accounting and market information as we describe in Section 4.

2.2 Covariance Decomposition

This section describes how we decompose the realized covariance matrix of returns for all assets in two components: factor covariance matrix and residual covariance matrix. Let $\boldsymbol{\Sigma}_t$ denote the realized covariance matrix of returns at time t, that is, $\boldsymbol{\Sigma}_t = \text{cov}(r_t)$. By using equation 3 and the assumption $\mathbb{E}(\boldsymbol{\varepsilon}_t|\boldsymbol{f}_t) = \mathbf{0}$, we have:

$$\boldsymbol{\Sigma}_t = \text{cov}(\boldsymbol{B}_t'\boldsymbol{f}_t) + \text{cov}(\boldsymbol{\varepsilon}_t) = \boldsymbol{B}_t'\boldsymbol{\Sigma}_{f,t}\boldsymbol{B}_t + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon},t} \tag{5}$$

where $\boldsymbol{B}_t$ is a $K \times N$ matrix of loadings of N assets on K factors, $\boldsymbol{\Sigma}_{f,t}$ is the $K \times K$ factor covariance matrix and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon},t}$ is the $N \times N$ residual covariance matrix, all in time t. Since each factor is a linear combination of returns, the factor covariance matrices can be obtained by using equation 4 and the known values of $\boldsymbol{\Sigma}_t$, that is

$$\boldsymbol{\Sigma}_{f,t} = \text{cov}(\boldsymbol{f}_t) = \text{cov}(\boldsymbol{W}'\boldsymbol{r}_t) = \boldsymbol{W}'\boldsymbol{\Sigma}_t\boldsymbol{W} \tag{6}$$

The factor loadings $\boldsymbol{B}_t$ are calculated using a similar procedure (see Appendix A) and the values of $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon},t}$ are simply given by the difference $\boldsymbol{\Sigma}_t - \boldsymbol{B}_t'\boldsymbol{\Sigma}_{f,t}\boldsymbol{B}_t$. It is common to assume that $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon},t}$ is diagonal, but we will be less restrictive in this work.

3 Forecasting Methodology

With the decomposition achieved in equation 5, we forecast the complete covariance matrix $\boldsymbol{\Sigma}_t$ by plugging separate forecasts of $\boldsymbol{B}_t$, $\boldsymbol{\Sigma}_{f,t}$, and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon},t}$. That is,

$$\widehat{\boldsymbol{\Sigma}}_{t+1|t} = \widehat{\boldsymbol{B}}'_{t+1|t}\widehat{\boldsymbol{\Sigma}}_{f,t+1|t}\widehat{\boldsymbol{B}}_{t+1|t} + \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon},t+1|t} \tag{7}$$

The next subsections describe how each of the individual components are forecast.

3.1 Forecasting $\boldsymbol{\Sigma}_{f,t}$

Since the number of factors is much smaller than the number of assets, one could propose the use of an AR model for the factor covariance matrix dynamics, as in equation (1). Despite the reduction in dimensionality achieved by using a factor model, the number of parameters in this configuration is still quite large (note that each equation would have $K(K+1)/2+1$ parameters). To reduce this number to a more manageable one, we follow the heterogeneous autoregressive (HAR) model proposed by Corsi (2009). In this model, the predictors are obtained from simple average of daily data, computed for different horizons (daily, weekly, and monthly). In our case, daily, weekly, and realized covariance matrices of factors are given by:

$$\begin{aligned} \boldsymbol{\Sigma}^{day}_{f,t} &= \boldsymbol{\Sigma}_{f,t} \\ \boldsymbol{\Sigma}^{week}_{f,t} &= \tfrac{1}{5}(\boldsymbol{\Sigma}_{f,t} + \boldsymbol{\Sigma}_{f,t-1} + \cdots + \boldsymbol{\Sigma}_{f,t-4}) \\ \boldsymbol{\Sigma}^{month}_{f,t} &= \tfrac{1}{22}(\boldsymbol{\Sigma}_{f,t} + \boldsymbol{\Sigma}_{f,t-1} + \cdots + \boldsymbol{\Sigma}_{f,t-21}) \end{aligned} \tag{8}$$

Using the notation $\boldsymbol{y}_{f,t} = \text{vech}(\boldsymbol{\Sigma}_{f,t})$, $\boldsymbol{y}^{day}_{f,t} = \text{vech}(\boldsymbol{\Sigma}^{day}_{f,t})$, $\boldsymbol{y}^{week}_{f,t} = \text{vech}(\boldsymbol{\Sigma}^{week}_{f,t})$, and $\boldsymbol{y}^{month}_{f,t} = \text{vech}(\boldsymbol{\Sigma}^{month}_{f,t})$, the dynamic process for $y_{f,t}$ is given by:

$$\boldsymbol{y}_{f,t} = \boldsymbol{\omega} + \boldsymbol{\Phi}_{day}\boldsymbol{y}^{day}_{f,t-1} + \boldsymbol{\Phi}_{week}\boldsymbol{y}^{week}_{f,t-1} + \boldsymbol{\Phi}_{month}\boldsymbol{y}^{month}_{f,t-1} + \boldsymbol{\epsilon}_t \tag{9}$$

where $\mathbf{\Phi}_{day}$, $\mathbf{\Phi}_{week}$, and $\mathbf{\Phi}_{month}$ are $M \times M$ matrices, where $M = K(K+1)/2$ is the number of unique entries on the the factor covariance matrix. $\boldsymbol{\omega}$ is a $M \times 1$ vector of intercepts.

3.1.1 LASSO

Due the high number of parameters in equation (9), direct estimation with ordinary least squares (OLS) could result in overfitting, harming the precision of the model forecasts. LASSO shrinks these estimates by imposing a penalty related to the magnitude of the coefficients. This estimation effectively sets most estimates to zero (Tibshirani (1996)). This methodology has been shown to provide a higher out-of-sample forecasting precision, while the reduced number of predictors makes the interpretation of the model easier.

We estimate (9) equation by equation. Consider a sample size of T and let $_t = (1, \boldsymbol{y}_{f,t-1}^{day}{}', \boldsymbol{y}_{f,t-1}^{week}{}', \boldsymbol{y}_{f,t-1}^{month}{}')'$ be the $3m+1$ vector of explanatory variables and $\boldsymbol{Z} = (\boldsymbol{Z}_T, \ldots, \boldsymbol{Z}_1)'$ the $T \times (3m+1)$ matrix of covariates. Let $\boldsymbol{y}_{f,i} = (y_{f,T,i}, \ldots, y_{f,1,i})'$ be the $T \times 1$ vector of observations on the ith equation of (9), $i = 1, \ldots, M$, and $\boldsymbol{\epsilon}_{f,i} = (\epsilon_{f,T,i}, \ldots, \epsilon_{f,1,i})'$ the corresponding vector of error terms. With $\boldsymbol{\gamma}_{f,i} = (\omega_i, \boldsymbol{\beta}_i')'$ being the $3M+1$ vector of true parameters for equation i, one can write:

$$\boldsymbol{y}_{f,i} = \boldsymbol{Z}\boldsymbol{\gamma}_{f,i} + \boldsymbol{\epsilon}_{f,i}, \qquad i = 1, \ldots, M \tag{10}$$

Each vector $\boldsymbol{\gamma}_{f,i}$ is then estimated by minimizing the following objective function:

$$L(\boldsymbol{\gamma}_{f,i}) = \frac{1}{T}||\boldsymbol{y}_{f,i} - \boldsymbol{Z}\boldsymbol{\gamma}_{f,i}||^2 + 2\lambda_T||\beta_i||_{\ell 1} \tag{11}$$

The penalty parameter λ_T determines how much penalization is imposed on the size of the coefficients. In our setup, the value of λ_T is determined by minimizing the Bayesian information criterion (BIC). For equation i and penalty parameter λ, the BIC is given by:

$$BIC_i(\lambda) = T \times log(\widehat{\boldsymbol{\epsilon}}'_{\lambda,i}\widehat{\boldsymbol{\epsilon}}_{\lambda,i}) + \sum_{j=1}^{3M} \mathbb{1}(\widehat{\beta}_{ij}^{\lambda} \neq 0)\log(T) \tag{12}$$

where $\widehat{\epsilon}_{\lambda,i}$ is the estimated vector of error terms for penalty λ. After obtaining $\widehat{\boldsymbol{\gamma}}_{f,i}$ from equation (11), the one-step-ahead forecast is given by

$$\widehat{y}_{f,T+1,i} = \widehat{\gamma}'_{f,i} Z_T, \quad i = 1, \dots, M, \tag{13}$$

which can be used to provide $\widehat{\boldsymbol{\Sigma}}_{f,T+1}$.

3.2 Forecasting $\boldsymbol{B}_t$

Instead of using unconditional loadings on factors, we assume that betas change daily. In Appendix B, we show the distribution of the fractional integration parameter for the beta series. The results show that these series present high persistence, similar to what is observed on realized covariance data. Based on this evidence, and to maintain consistency with the rest of our methodology, we forecast each series of betas using HAR models. We model each beta individually and estimate the equations via OLS.

3.3 Forecasting $\boldsymbol{\Sigma}_{\varepsilon,t}$

Since the residual covariance matrix dimension is $N \times N$, the curse of dimensionality remains a concern when forecasting $\boldsymbol{\Sigma}_{\varepsilon,t}$. A factor model may not explain all the covariances between assets depending on the number and choice of factors, implying the the covariance matrix of the residuals is not diagonal. Instead of imposing the common hypothesis that $\boldsymbol{\Sigma}_{\varepsilon,t}$ is diagonal (as in Fan, Fan, and Lv (2008)), we consider a less restrictive assumption that stocks in the same sector may still co-vary even after controlling for the standard factors used in the finance literature.

We assume that $\boldsymbol{\Sigma}_{\varepsilon,t}$ is block-diagonal where blocks are defined by industry classification. We assume that there is no contemporaneous correlation between assets in different sectors after controlling for factor exposure. We argue that this assumption is not too strong in Section 4.4, where we analyze residual correlations between stocks in different sectors.

We assume that each asset belongs to an industrial sector s, where S is the total number of sectors and $S << N$. Without loss of generality, we can order assets so that the residual covariance matrix has sector covariance matrices near its diagonal:

$$\mathbf{\Sigma}_{\boldsymbol{\varepsilon},t} = \begin{pmatrix} \mathbf{\Sigma}^1_{\boldsymbol{\varepsilon},t} & & \\ & \ddots & \\ & & \mathbf{\Sigma}^S_{\boldsymbol{\epsilon},t} \end{pmatrix} \tag{14}$$

Depending on the selected industry classification, the number of assets on each group s (call it N^s) can still be quite large as each block has $M^s = N^s(N^s+1)/2$ unique elements.

Our second assumption is that the dynamics of each block $\mathbf{\Sigma}^s_{\boldsymbol{\varepsilon},t}$ depends only on the t-1 elements of the same block, i.e., only $\mathbf{\Sigma}^s_{\boldsymbol{\varepsilon},t-1}$ predict values in $\mathbf{\Sigma}^s_{\boldsymbol{\varepsilon},t}$. In order to further simplify this model, we also consider an additional restriction that only past variances matter in this prediction. This last assumption relies on the previous evidence in Callot, Kock, and Medeiros (2017) that past variances are more frequently selected by Lasso as good predictors of covariance and variance terms.

With the notation $\boldsymbol{y}^s_{\epsilon,t} = \text{vech}(\mathbf{\Sigma}^s_{\boldsymbol{\varepsilon},t})$, we have:

$$\boldsymbol{y}^s_{\boldsymbol{\varepsilon},t} = \boldsymbol{\omega}^s_{\epsilon} + \mathbf{\Phi}^s \mathbf{\Lambda}^s_{\boldsymbol{\varepsilon},t-1} + \boldsymbol{u}^s_{\boldsymbol{\varepsilon},t}, \qquad s = 1, \ldots, S, \tag{15}$$

where $\boldsymbol{\omega}^s_{\boldsymbol{\varepsilon}}$ is a $M^s \times 1$ vector of intercepts, $\mathbf{\Phi}^s$ is a $M^s \times N^s$ matrix of coefficients, $\mathbf{\Lambda}^s_{\boldsymbol{\varepsilon},t-1} = \mathsf{diag}(\mathbf{\Sigma}^s_{\boldsymbol{\varepsilon},t-1})$ is a $N^s \times 1$ vector of past variances, and $\boldsymbol{u}^s_{\boldsymbol{\varepsilon},t}$ is the vector of errors.

The parameters estimation is done block by block, that is, each equation in (15) is estimated separately. The procedure is the same as the one used for the factor covariance matrix model: LASSO regression equation by equation, using a sample size of T. We then regroup the one-day ahead forecast for each group, $\widehat{\mathbf{\Sigma}}^s_{\boldsymbol{\varepsilon},T+1}$, to form the full residual covariance matrix forecast $\widehat{\mathbf{\Sigma}}_{\boldsymbol{\varepsilon},T+1}$.

4 Data and Descriptive Analysis

4.1 Realized Return Covariance Matrices

The data consists of daily realized covariance matrices of returns for constituents of the S&P 500 index. These matrices were constructed using 5-minutes returns by the composite realized kernel method (discussed in Lunde, Shephard, and Sheppard (2016) and provided by the authors). The full sample comprises all business days between January 2006 and December 2011. We consider companies that remained in the index and had balance sheet data available for the full sample period, resulting in a total of 430 stocks. With these considerations, the data set consists of 1495 daily 430×430 realized covariance matrices of returns.

4.2 Factors

We construct each factor as the return time series for a long-short stock portfolio derived from individual sorts of the underlying stocks on different signals. Since our approach uses the loading matrix W as an input to calculate factor covariance matrices and factor loadings, we could not use the widely available data on financial factors series of returns (such as in Kenneth French website). Instead, we had to rank our own universe of stocks in different signals and calculate the matrices W for our sample.

In total, we use 7 factors that have been widely used in the finance literature. Besides the market factor, we also consider: Size (SMB) and Value (HML) (Fama and French 1993), Gross Profitability (Novy-Marx 2013), Investment (Lyandres, Sun, and Zhang 2008), Asset Growth (Cooper, Gulen, and Schill 2008), and Accruals (Sloan 1996). We report a detailed description of each factor construction in Appendix C. We use three different combinations of factors with 3, 5, and 7 factors. They are denoted by 3F (Market, Size, and Value), 5F (3F + Gross Profitability and Investment), and 7F (5F

+ Asset Growth and Accruals), respectively.

4.3 Sector Classification

Each stock is classified in one of 10 sectors, following the Standard Industrial Classification (SIC). Table 1 shows the number of companies from our sample on each sector. Notice that some groups are quite large (the group Others, for instance, has more than 100 stocks). This motivates the use of additional restrictions we discussed in subsection 3.3.

Table 1: Number of Stocks per Sector

Sector	Number of Stocks
Consumer Non-Durables	31
Consumer Durables	8
Manufacturing	65
Oil, Gas, and Coal Extraction	32
Business Equipment	61
Telecommunications	10
Wholesale and Retail	45
Healthcare, Medical Equipments, and Drugs	26
Utilities	36
Others	116

4.4 Residual Covariance Matrices

In this section, we analyze the series of residual covariance matrices for the 3 factors combinations. Using notation from previous sections, these series are given by $\Sigma_{\epsilon,t} = \Sigma_t - B_t' \Sigma_{f,t} B_t$, for each of the 1495 days in the sample. To analyze if these matrices are approximately block diagonal, we follow the procedure discussed in Ait-Sahalia and Xiu (forthcoming). First, we transform the covariance matrices in correlation matrices. We classify a correlation as significant if it is higher than 0.15 in at least 1/3 of the sample. We then plot the significant relations as dots, while the rest of the points are left in blank. Sectors in Table 1 are represented as red squares (in the same order) . Figure 1 shows the results when we apply the criterion to the full covariance matrices Σ_t, before the factor decomposition. We can see that most correlations are classified as significant in the original data.

Figure 1: Significant Correlations Between Stocks on the Realized Covariance Matrices

The blue dots represent the correlations between stocks that are higher than 0.15 in at least 1/3 of the sample days. Red squares represent the groups defined by sector industrial classification (SIC). The axis have indexes that correspond to the 430 stocks in our sample. This plot is calculated using the complete series of realized covariance matrices, before any factor decomposition.

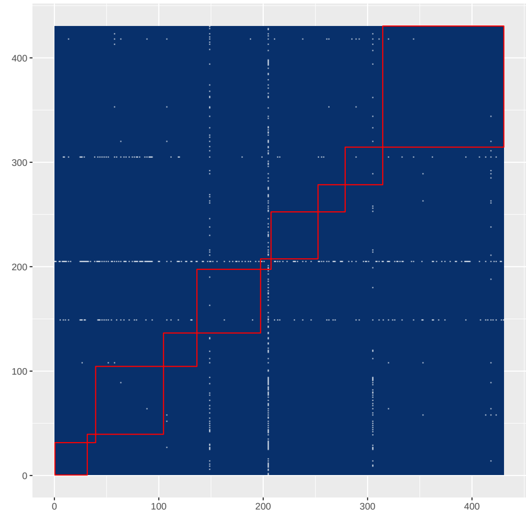
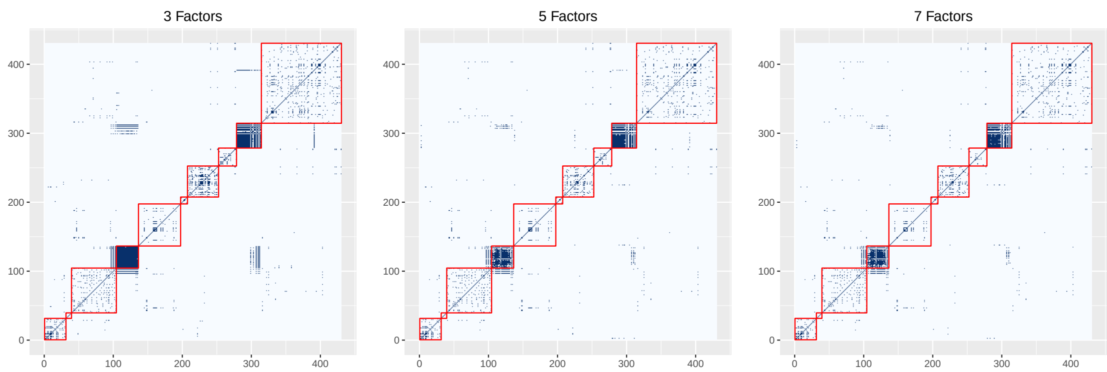


Figure 2: Significant Correlations Between Stocks on the Residual Covariance Matrices

The blue dots represent the correlations between stocks that are higher than 0.15 in at least 1/3 of the sample days. Red squares represent the groups defined by sector industrial classification (SIC). The axis have indexes that correspond to the 430 stocks in our sample. This plot show the results for the 3 series of residual covariance matrices obtained after using the factor decomposition, for three factors configurations.



When we follow this procedure for the residual covariance matrices (Figure 2), we obtain plots that are much more sparse than the previous. This means that most of the correlations are not significant according to the criterion. Besides, we can see that most of the dots are contained inside the blocks, meaning that the majority of significant correlations are between stocks in the same sector. We interpret the three plots as evidence that we would not lose much information by assuming block diagonality for the residual covariance matrices. Notice that the results are robust across different factors configurations.

5 Forecasting Results

This section reports the forecasting results using the methodology described in Section 3. Subsection 5.1 shows the forecasting results for Σ_f, while subsection 5.2 shows the results for the complete covariance matrix Σ. In both cases, daily forecasts are computed using a sample size of 1000 observations. Since we need 22 days to compute the first monthly regressor, $\Sigma_{f,t}^{month}$, our out-of-sample forecasts comprise days 1023 to 1495 (T_1 to T_2), totaling 473 daily forecasts.

We evaluate our forecasts by using the ℓ_2-norm for the vector of errors, that is, $||\hat{e}_{T+1}|| = ||vech(\hat{\Sigma}_{T+1} - \Sigma_{T+1})||$. We compare different methods by using the average ℓ_2-forecast error:

$$\text{average } \ell_2\text{-forecast error} = \frac{1}{T_2 - T_1 + 1} \sum_{T=T_1}^{T=T_2} ||\hat{e}_{T+1}|| \tag{16}$$

In all cases, our benchmark forecast is a random walk model: $\hat{\Sigma}_{f,T+1} = \Sigma_{f,T}$ and $\hat{\Sigma}_{T+1} = \Sigma_T$ for the complete covariance matrix.

5.1 Factor Covariance Matrix

Table 2 shows the forecast results for the factor covariance matrices, following the method described in Section 3.1 (we refer to this method as FHAR, from Factor HAR, hereafter). Results are uniformly stronger when we apply a log-matrix transformation as proposed by Chiu, Leonard, and Tsui (1996). Prior to estimation, we apply the matrix logarithm to all data, that is, $\Omega_{f,t} = \log(\Sigma_{f,t})$. We then use the method FHAR with $\Omega_{f,t}$ to construct $\hat{\Omega}_{f,t+1}$. Finally, we revert the transformation by applying the matrix exponential and get our forecasts, that is, $\hat{\Sigma}_{f,t+1} = \exp(\hat{\Omega}_{f,t+1})$. We also report results without applying the log-matrix transformation.

Table 2: Forecast Precision for Factor Covariance Matrices

	ℓ_2	ℓ_2 / $\ell_{2,benchmark}$	
Model	Benchmark	FHAR	FHAR, Log-matrix
3F	0.44	0.98	0.90
5F	0.51	0.95	0.89
7F	0.62	0.99	0.86

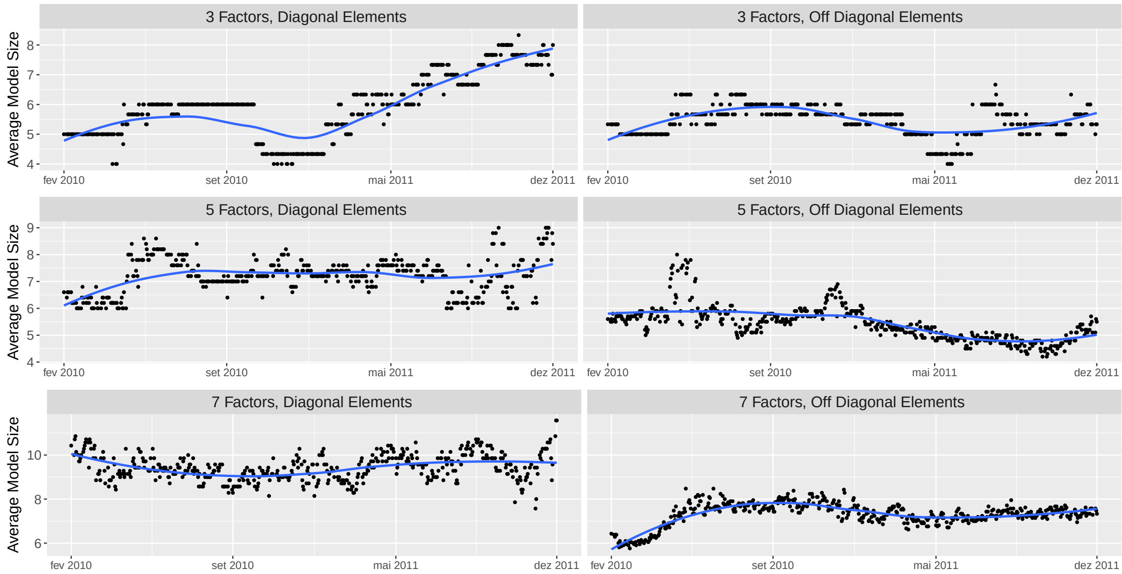
Notes: ℓ_2 represents the average ℓ_2-forecast error, calculated for 473 days. $\ell_2/\ell_{2,benchmark}$ represents the ratio of average error for other methods relative to the benchmark value. FHAR stands for the methodology described in Section 3.1. Each line represents a factor configuration.

From Table 2, we see that using the log matrix transformation considerably improves the forecasts. After this transformation, variance and covariance series become smoother than the original data. This reduces the weights of outliers when fitting the model and improves forecasting precision. Another advantage is that the exponential matrix is positive definite by construction. This consideration is important, since we will need the inverse of the covariance matrix Σ when solving the minimum variance problems in Section 6.

Since we estimate all equations daily, we now investigate their evolution over time. Figure 3 shows the average number of variables selected by the Lasso over the 473 days, for the FHAR Log-matrix models. To illustrate, consider the upper-left panel on this figure. It shows the average model size for the diagonal equations in the FHAR Log-matrix model with 3 factors. In other words, it displays the mean size for the equations of σ^2_{market}, σ^2_{SMB}, and σ^2_{HML}. Panels on the right show the same results for the covariance equations. From top to bottom, we vary the number of factors.

Figure 3: Average Number of Selected Variables in FHAR Log-matrix models

These panels shows the daily average number of variables selected by the lasso for variance equations and covariance equations. Averages are simple means of the number of variables selected in each class of equations on a given day. We sort the panels by number of factors in the model (increasing from top to bottom) and by variable class (variances on the left and covariances on the right). Blue lines are local polynomial regressions. The maximum number predictive variables for each configuration are: 18 (3 factors), 45 (5 factors), and 76 (7 factors).



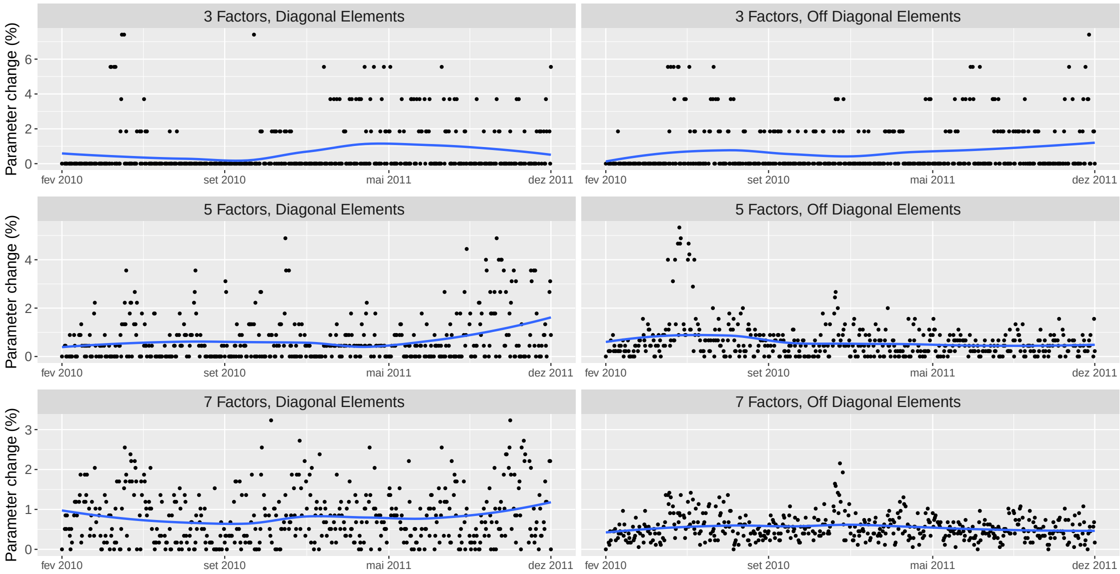
In all cases, the Lasso reduces the number of selected variables substantially. This reduction is most noticeable for the FHAR Log-matrix model with 7 factors, in which the number of selected variables fluctuates around 10 (out of 76 potential predictors) for the variance equations and even less for the covariance equations.

Another interesting feature is that average equation sizes are stable over time. Figure 4 shows the average change on the number of selected variables from one day to the next. We consider parameters changing from zero to non-zero values or the opposite direction. All results are reported in percentage (relative to the maximum number of variables on each equation). For the 3 factors models, both variance and covariance equations remain unchanged in most days. Despite the stronger variation on the 5 and 7 factors configurations, the percentage changes are smaller than 5% in almost all days. In general, variance equations are slightly more stable.[2]

[2]We present results for the average equation size and average equation change for the residual covariance matrices forecasts in Appendix D.

Figure 4: Average Change on the Number of Selected Variables in FHAR Log-matrix models

These panels shows the daily average change in the number of variables selected by the lasso for variance equations and covariance equations. We count a change when a variable goes from being zero to non-zero or from being non-zero to zero. Results are presented as percentage relative to the maximum number of variables on each equation. We sort the panels by number of factors in the model (increasing from top to bottom) and by variable class (variances on the left and covariances on the right). Blue lines are local polynomial regressions.



5.2 Complete Covariance Matrix

Table 3 shows the forecast results for the complete covariance matrix (430 stocks). In this case, VHAR denotes the complete methodology described in Section 3. We also show the results when we apply the Log-matrix transformation for both the factor covariance matrix and residual covariance matrix forecasts. Table 3 also reports three alternative models for comparison purposes:

1. Exponentially weighted moving average (EWMA) with smoothing parameter $\lambda = 0.96$;
2. A Dynamic Conditional Correlation model and a BEKK model, both with nonlinear shrinkage, as proposed by Engle, Ledoit, and Wolf (2017). These approaches will be called DCC-NL and BEKK-NL hereafter.

Different from our methodology, the alternative models use daily returns data, instead of using realized covariance matrices as inputs. We compare our results to these models to see if the use of a richer dataset (remember that realized covariances are calculated from high-frequency data) brings gains in terms of portfolio performance and economic value.

Table 3: Forecast Precision for Complete Covariance Matrices

Return Models	$\ell_2/\ell_{2,b}$	VHAR	$\ell_2/\ell_{2,b}$	VHAR, Log-matrix	$\ell_2/\ell_{2,b}$
EWMA	6.93	3F	0.92	3F	0.89
BEKK-NL	1.71	5F	0.91	5F	0.89
DCC-NL	1.71	7F	0.90	7F	0.89
Benchmark error ($\ell_{2,b}$)			341.57		

Notes: ℓ_2 represents the average ℓ_2-forecast error, calculated for 473 days. $\ell_{2,b}$ is this error calculated using the benchmark. $\ell_2/\ell_{2,benchmark}$ represents the ratio between these values. The column "Return Models" has the models based on daily return data. The columns "VHAR" and "VHAR, Log-matrix" have our results (without and with log-matrix transformation, respectively). 3F, 5F, and 7F stand for the 3 factors, 5 factors, and 7 factors configurations for the factor covariance matrix.

The first observation on the results in Table 3 is that there is only a single average ℓ_2-error value for the benchmark, EWMA, BEKK-NL and DCC-NL. This is different from table 2 since the aforementioned models do not apply a factor decomposition and hence do not depend on the number of factors as in our case. Instead, we use them to forecast $\hat{\Sigma}_{T+1}$ directly.

Models based on daily returns do not outperform the benchmark in any of the cases, with a particular bad performance for the EWMA. Despite having performed much better, BEKK-NL and DCC-NL have errors that are almost 2 times larger than the benchmark. Our methodology is able to beat the benchmark results for all factors configurations, with a gain of 10% in forecast precision for the 7F VHAR model. As before, results are improved when we use the log matrix transformation.

6 Portfolio Selection

Despite the encouraging results when forecasting covariance matrices, one could argue that our measure of forecast precision is too aggregate, specially when taking into account that we have more than 90000 unique entries being forecast simultaneously. To evaluate the economic advantages provided by our forecasts, we use them to construct investment portfolios. Besides analyzing the traditional portfolio statistics, such as returns, standard deviations and Sharpe ratios, we also use the approach of Fleming, Kirby, and Ostdiek (2003) to evaluate the economic value of our forecasts.

6.1 Global Minimum Variance Portfolios

Consider the problem of an investor at time $t = t_0, ..., T-1$ who wishes to construct a minimum variance portfolio to be held in time $t+1$. For this minimization problem, the investor needs to forecast the future covariance matrix, $\hat{\Sigma}_{t+1}$. The optimization problem consists in choosing a vector of weights $\hat{w}_{t+1}$ (dimension $N \times 1$):

$$\begin{aligned} \hat{w}_{t+1} = \arg\min_{w_{t+1}} \quad & w'_{t+1}\hat{\Sigma}_{t+1}w_{t+1} \\ \text{subject to} \quad & w'_{t+1}\mathbf{1} = 1 \end{aligned} \tag{17}$$

We use our VHAR method with log matrix transformation, since they provide the best results in terms of forecasting precision. We evaluate this model against the random walk benchmark as before, and also compare it to EWMA, BEKK-NL and DCC-NL. We evaluate ex-post portfolio performance, that is, we use our time t estimated weights, $\hat{w}_{t+1}$, with data from t+1. In the following, $\hat{w}_{it}$ is the i-th component of $\hat{w}_t$, We show the following statistics:

1. Max. weight: $\max_{t_0+1 \le t \le T} \max_{1 \le i \le N}(\hat{w}_{it})$ for $t = t_0+1, ..., T$ and $i = 1, ..., N$.
2. Min. weight: $\min_{t_0+1 \le t \le T} \min_{1 \le i \le N}(\hat{w}_{it})$ for $t = t_0+1, ..., T$ and $i = 1, ..., N$.
3. Proportion of leverage: $\frac{1}{n(T-t_0)} \sum_{t=t_0+1}^{T} \sum_{i=1}^{N} I(\hat{w}_{it} < 0)$.
4. Average turnover: $\frac{1}{n(T-t_0)} \sum_{t=t_0+1}^{T} \sum_{i=1}^{N} |\hat{w}_{it} - \hat{w}_{it}^{hold}|$, where $\hat{w}_{it}^{hold} = \hat{w}_{it-1}\frac{1+r_{it-1}}{1+r_{pt-1}}$. r_{pt} is the portfolio return at time t, r_{it} is the stock i return at time t, and $\hat{w}_{it}^{hold}$is the weight of stock i on the hold portfolio. The hold portfolio at time t+1 is defined as the resulting portfolio from keeping all the stocks from period t.

5. Average excess return: $\mu_p^e = \frac{1}{(T-t_0)} \sum_{t=t_0+1}^{T} (r_{pt}^e) = \frac{1}{(T-t_0)} \sum_{t=t_0+1}^{T} (\hat{w}_t' r_t - r_{f,t})$, where $r_{f,t}$ is the risk free rate
6. Cumulative Return: $\prod_{t=t_0+1}^{T} (1 + r_{pt})$.
7. Standard deviation: $\sigma_p = \sqrt{\frac{1}{(T-t_0)} \sum_{t_0+1}^{T} (r_{pt} - \frac{1}{(T-t_0)} \sum_{t=t_0+1}^{T} r_{pt})^2}$.
8. Sharpe ratio: $\frac{\mu_p^e}{\sigma_p}$.
9. Average diversification ratio: $\frac{1}{(T-t_0)} \sum_{t_0+1}^{T} \frac{\sum_{i=1}^{N} \hat{w}_{it} \sigma_{it}}{\sigma_{pt}}$, where $\sigma_{pt} = \hat{w}_t' \Sigma_t \hat{w}_t$.

To complement our analysis, we also use the methodology described in Fleming, Kirby, and Ostdiek (2003) to evaluate the economic value of our method when compared against the benchmark, EWMA, BEKK-NL and DCC-NL. This analysis assumes an investor with utility given by:

$$U(r_{pt}) = (1 + r_{pt}) - \frac{\gamma}{2(1+\gamma)(1 + r_{pt})^2} \tag{18}$$

where γ is the investor's risk aversion coefficient. The variable economic value is then defined as the value Δ such that, for different portfolios p_1 and p_2, we have $\sum_{t=t_0+1}^{T} U(r_{p_1 t}) = \sum_{t=t_0+1}^{T} U(r_{p_2 t} - \Delta)$.

Table 4 shows these results for the minimum variance portfolio optimization problem, presented in equation 17. In terms of standard deviations and Sharpe ratios, the VHAR class of models perform better. When compared to the random walk benchmark, for instance, our models are able to practically double the Sharpe ratio and reduce standard deviation in at least 60%. EWMA has, by far, the worst performance, with the second highest standard deviation and the lowest Sharpe ratio. The BEKK-NL and DCC-NL models have very good results, specially when considering that they rely solely on returns data to forecast covariance matrices.

In terms of average excess returns and cumulative returns, our models have a slightly inferior performance than the others (except for the EWMA). Since the optimization problem focus on minimizing the portfolio variance, it is not obvious that we should achieve a higher return as this is not the objective of this portfolio. For this reason, we do not see the difference in returns as a problem for our methodology. Moreover, the small difference is more than compensated by the substantial reduction in standard deviation (as evidenced by the higher Sharpe ratios).

We can relate the economic values to the patterns observed in returns and standard deviations. For representative investors with low risk aversion ($\gamma = 1$, we see that most economic values are negative (except for the EWMA). This tendency is reversed for high risk aversion investors ($\gamma = 10$, when our models have positive economic value over all the competing alternatives. For moderately risk averse investors ($\gamma = 5$), our models have economic value over the benchmark and the EWMA model. In general, BEKK-NL and DCC-NL are more balanced in terms of the risk-return trade-off and therefore harder to beat.

Table 4: Statistics for Daily Portfolios - Global Minimum Variance

	Benchmark	EWMA	BEKK-NL	DCC-NL	3-Factor VHAR (Log matrix)	5-Factor VHAR (Log matrix)	7-Factor VHAR (Log matrix)
Max. Weight	0.13	0.51	0.08	0.30	0.47	0.47	0.49
Min.Weight	-0.17	-0.46	-0.05	-0.06	-0.04	-0.04	-0.05
Proportion of leverage (%)	44.30	49.17	45.11	51.73	44.88	44.89	45.26
Average Turnover (%)	1.80	0.27	0.11	0.21	0.20	0.19	0.20
Average Excess Return (%)	20.25	5.33	16.58	19.78	14.60	14.65	14.24
Cumulative Return (%)	39.96	7.58	34.90	42.07	30.77	30.99	29.98
Standard Deviation (%)	23.41	17.72	12.24	15.73	9.20	8.70	8.74
Sharpe Ratio	0.87	0.30	1.35	1.26	1.59	1.68	1.63
Diversification Ratio	3.29	1.01	3.03	3.53	5.02	4.87	4.96
Economic Value ($\gamma = 1$)							
Benchmark	-	-	-	-	-3.33	-3.24	-3.65
EWMA	-	-	-	-	10.42	10.51	10.10
BEKK-NL	-	-	-	-	-1.65	-1.55	-1.97
DCC - NL	-	-	-	-	-4.37	-4.27	-4.69
Economic Value ($\gamma = 5$)							
Benchmark	-	-	-	-	5.95	6.22	5.79
EWMA	-	-	-	-	15.00	15.27	14.84
BEKK-NL	-	-	-	-	-0.34	-0.07	-0.50
DCC - NL	-	-	-	-	-1.10	-0.83	-1.26
Economic Value ($\gamma = 10$)							
Benchmark	-	-	-	-	17.56	18.05	17.60
EWMA	-	-	-	-	20.71	21.21	20.76
BEKK-NL	-	-	-	-	1.30	1.80	1.35
DCC - NL	-	-	-	-	3.00	3.49	3.05

Table 4 shows that some portfolios have extreme short positions. For all cases, the proportion of leverage is close to 50%. Since shorting stocks may not be feasible or may be potentially costly, we consider these positions potentially extreme. Values for maximum and minimum weights are also quite large in some cases. In one of the cases, more than 50% of the portfolio is allocated into a single stock. We resort to what we consider a more realistic investor problem in the next subsection.

6.2 Restricted Minimum Variance Portfolios

In this section, we solve a problem similar to equation 17, except that now we impose two additional restrictions. First, we allow maximum leverage to be 30% (in some sense, consistent with a 130-30 fund concept in the mutual fund industry). Second, we restrict the maximum weights on individual stocks to be 20% (in absolute value). The problem for an investor at time $t = t_0, ..., T-1$ is then given by:

$$\begin{gathered} \hat{w}_{t+1} = \underset{w_{t+1}}{\arg\min} \quad w'_{t+1}\hat{\Sigma}_{t+1}w_{t+1} \\ \text{subject to} \quad w'_{t+1}\mathbf{1} = 1, \\ \sum_{i=1}^{N} |w_{it+1}| I(w_{it} < 0) \leq 0.30 \quad \text{and} \quad |w_{it+1}| \leq 0.20 \end{gathered} \tag{19}$$

In Table 5 we report the same performance statistics as in last subsection. Again, our methodology generates better Sharpe ratios and lower standard deviations than the competing models. This time, we also obtain a better performance in terms of average returns and cumulative returns (except for the 5-factor model, which performed as well as the DCC-NL). We also see that the restrictions on maximum and minimum weights are binding in all cases, while the restriction on maximum daily leverage drastically reduces the proportion of leverage.

In this case, all models deliver positive economic values, regardless the level of risk aversion. This is due to the increase in average returns, while maintaining low standard deviations. On this setting, a representative investor would value our methodology by up to 10% when compared to the BEKK-NL.

Table 5: Statistics for Daily Portfolios - Restricted Minimum Variance

	Benchmark	EWMA	BEKK-NL	DCC-NL	3-Factor VHAR (Log matrix)	5-Factor VHAR (Log matrix)	7-Factor VHAR (Log matrix)
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
Proportion of leverage (%)	1.91	0.71	0.85	1.41	2.37	2.43	2.27
Average Turnover (%)	0.43	0.09	0.10	0.11	0.24	0.23	0.22
Average Excess Return (%)	12.58	10.11	10.07	13.93	15.51	13.88	16.03
Cumulative Return (%)	24.81	18.52	18.36	27.53	32.04	28.05	33.39
Standard Deviation (%)	13.26	15.25	15.49	14.71	12.76	12.76	12.59
Sharpe Ratio	0.95	0.66	0.65	0.95	1.22	1.09	1.27
Diversification Ratio	3.55	2.22	2.24	2.15	3.64	3.73	3.68
			Economic Value ($\gamma = 1$)				
Benchmark	-	-	-	-	3.00	1.37	3.55
EWMA	-	-	-	-	5.75	4.12	6.30
BEKK-NL	-	-	-	-	5.83	4.19	6.37
DCC-NL	-	-	-	-	1.85	0.22	2.40
			Economic Value ($\gamma = 5$)				
Benchmark	-	-	-	-	3.26	1.63	3.89
EWMA	-	-	-	-	7.15	5.52	7.78
BEKK-NL	-	-	-	-	7.37	5.73	8.00
DCC-NL	-	-	-	-	2.93	1.29	3.56
			Economic Value ($\gamma = 10$)				
Benchmark	-	-	-	-	3.59	1.96	4.33
EWMA	-	-	-	-	8.91	7.27	9.64
BEKK-NL	-	-	-	-	9.30	7.66	10.04
DCC-NL	-	-	-	-	4.28	2.64	5.02

6.3 Restricted Minimum Variance Portfolio with Target Return

We redo the calculations of last section with a target for the portfolio returns. Let $\hat{\mu}_{t+1}$ denote the expected N vector of returns at time t. In our case, we estimate these values as the moving average of 100 days. We set μ_{target}=10% and estimate:

$$
\begin{aligned}
\hat{w}_{t+1} = \underset{w_{t+1}}{\arg\min} \quad & w'_{t+1}\hat{\Sigma}_{t+1}w_{t+1} \\
\text{subject to} \quad & w'_{t+1}\hat{\mu}_{t+1} = \mu_{target}, \qquad w'_{t+1}\mathbf{1} = 1, \\
\sum_{i=1}^{N} |w_{it+1}| I(w_{it} < 0) \leq 0.30 \quad & \text{and} \quad |w_{it+1}| \leq 0.20
\end{aligned}
\tag{20}
$$

Table 6: Statistics for Daily Portfolios - Restricted Minimum Variance with Target for Returns

	Benchmark	EWMA	BEKK-NL	DCC-NL	3-Factor VHAR (Log matrix)	5-Factor VHAR (Log matrix)	7-Factor VHAR (Log matrix)
Max. Weight	0.20	0.20	0.20	0.20	0.20	0.20	0.20
Min.Weight	-0.19	-0.20	-0.20	-0.20	-0.20	-0.20	-0.19
Proportion of leverage (%)	7.89	5.76	7.71	7.55	8.63	8.85	8.99
Average Turnover (%)	0.45	0.18	0.19	0.18	0.26	0.26	0.25
Average Excess Return (%)	6.38	1.85	9.93	8.01	9.94	9.19	11.68
Cumulative Return (%)	11.08	1.54	18.25	14.28	18.86	17.09	22.88
Standard Deviation (%)	13.36	15.15	14.88	14.23	13.00	13.38	12.77
Sharpe Ratio	0.48	0.12	0.67	0.56	0.76	0.69	0.91
Diversification Ratio	3.80	2.33	2.41	2.31	3.90	3.92	3.90
Economic Value ($\gamma = 1$)							
Benchmark	-	-	-	-	3.61	2.81	5.38
EWMA	-	-	-	-	8.39	7.59	10.16
BEKK-NL	-	-	-	-	0.27	-0.53	2.04
DCC-NL	-	-	-	-	2.09	1.29	3.86
Economic Value ($\gamma = 5$)							
Benchmark	-	-	-	-	3.80	2.80	5.69
EWMA	-	-	-	-	9.59	8.60	11.48
BEKK-NL	-	-	-	-	1.32	0.32	3.21
DCC-NL	-	-	-	-	2.76	1.77	4.65
Economic Value ($\gamma = 10$)							
Benchmark	-	-	-	-	4.03	2.79	6.07
EWMA	-	-	-	-	11.10	9.86	13.14
BEKK-NL	-	-	-	-	2.63	1.39	4.67
DCC-NL	-	-	-	-	3.61	2.36	5.64

Like before, our models perform better in terms of standard deviation and Sharpe ratio. For average excess returns and accumulated returns, our models beat all the competing alternatives, except the BEKK, that have similar results to the 5-Factor VHAR model.

7 Conclusion

We propose a model to forecast very large realized covariance matrices of returns, applying it to the constituents of the S&P 500 on a daily basis. To deal with the curse of dimensionality, we decompose the return covariance matrix using standard firm-level factors (e.g. size, value, profitability) and use sectoral restrictions in the residual covariance matrix. This restricted model is then estimated using Vector Heterogeneous Autoregressive (VHAR) models estimated with the Least Absolute Shrinkage and Selection Operator (LASSO). Our methodology improves forecasting precision relative to standard benchmarks and leads to better estimates of the minimum variance portfolios.

Appendix

A Factor Loadings

This appendix describes how daily loadings on factors can be calculated using daily realised covariance matrices of returns Σ_t, daily factor covariance matrices $\Sigma_{f,t}$, and factor weights W (which are calculated yearly). To simplify the notation, we drop the subscript t and derive the equation for a daily matrix of loadings $\hat{B}$ (the final equation is applied to daily data, providing 1495 matrices $\hat{B}$). Letters with one subscript t represent vectors with cross-sectional data at time t, while letters with one superscript n represent time series data for asset n.

Assume that we have T observations for K factors and N assets (in our case, T can be thought as intra-daily high-frequency observations). Stacking all observations in matrix form,

$$F = \begin{pmatrix} f_{1,1} & \dots & f_{K,1} \\ \vdots & \ddots & \vdots \\ f_{1,T} & \dots & f_{K,T} \end{pmatrix} = \begin{pmatrix} f_1' \\ \vdots \\ f_T' \end{pmatrix}$$

$$R = \begin{pmatrix} r_{1,1} & \dots & r_{1,T} \\ \vdots & \ddots & \vdots \\ r_{N,1} & \dots & r_{N,T} \end{pmatrix} = \begin{pmatrix} r^{1\prime} \\ \vdots \\ r^{N\prime} \end{pmatrix} = \begin{pmatrix} r_1 & \dots & r_T \end{pmatrix} \tag{21}$$

From equation 4, we can rewrite the matrix F,

$$F = \begin{pmatrix} w_1' r_1 & \dots & w_K' r_1 \\ \vdots & \ddots & \vdots \\ w_1' r_T & \dots & w_K' r_T \end{pmatrix} = \begin{pmatrix} r_1' w_1 & \dots & r_1' w_K \\ \vdots & \ddots & \vdots \\ r_T' w_1 & \dots & r_T' w_K \end{pmatrix} = \begin{pmatrix} r_1' \\ \vdots \\ r_T' \end{pmatrix} \begin{pmatrix} w_1 & \dots & w_K \end{pmatrix} = R'W \tag{22}$$

With this setup, consider the linear model of asset n, n in (1,...,N), on K factors:

$$\underbrace{\begin{pmatrix} r_{n,1} \\ \vdots \\ r_{n,T} \end{pmatrix}}_{r^n} = \underbrace{\begin{pmatrix} f_1 \\ \vdots \\ f_T \end{pmatrix}}_{F} \underbrace{\begin{pmatrix} b_{n,1} \\ \vdots \\ b_{n,K} \end{pmatrix}}_{b_n} + \begin{pmatrix} \epsilon_{n,1} \\ \vdots \\ \epsilon_{n,T} \end{pmatrix} \tag{23}$$

The OLS estimator for b_n, $\hat{b}_n$, are given by:

$$\hat{b}_n = (F'F)^{-1}F'r^n \qquad \text{for i in 1,..,N} \tag{24}$$

In matrix form for all n,

$$\begin{aligned} \hat{B} = \begin{pmatrix} \hat{b}_1 & \dots & \hat{b}_N \end{pmatrix} &= (F'F)^{-1}F' \begin{pmatrix} r^1 & \dots & r^N \end{pmatrix} = \\ &= (F'F)^{-1}F' \begin{pmatrix} r'_1 \\ \vdots \\ r'_T \end{pmatrix} = (F'F)^{-1}F'R' = \\ &= (F'F)^{-1}W'RR' \xrightarrow{mean=0} (\Sigma_f)^{-1}W'\Sigma \end{aligned} \tag{25}$$

with equation 22 being used on the last line. Again, in our setup, $\hat{B}$ is different for every day in the sample.

B Betas Long Memory

Figure 5: Estimated Fractional Differencing Parameter (d) - Whittle Method

These panels show the distribution of fractional differencing parameter (d) estimated by the Whittle method. Each panel correspondes to one of the factors configuration (3, 5, or 7 factors). In total, there are $430 \times$ number of factors series of betas for each case. We plot the distribution of d for these series.

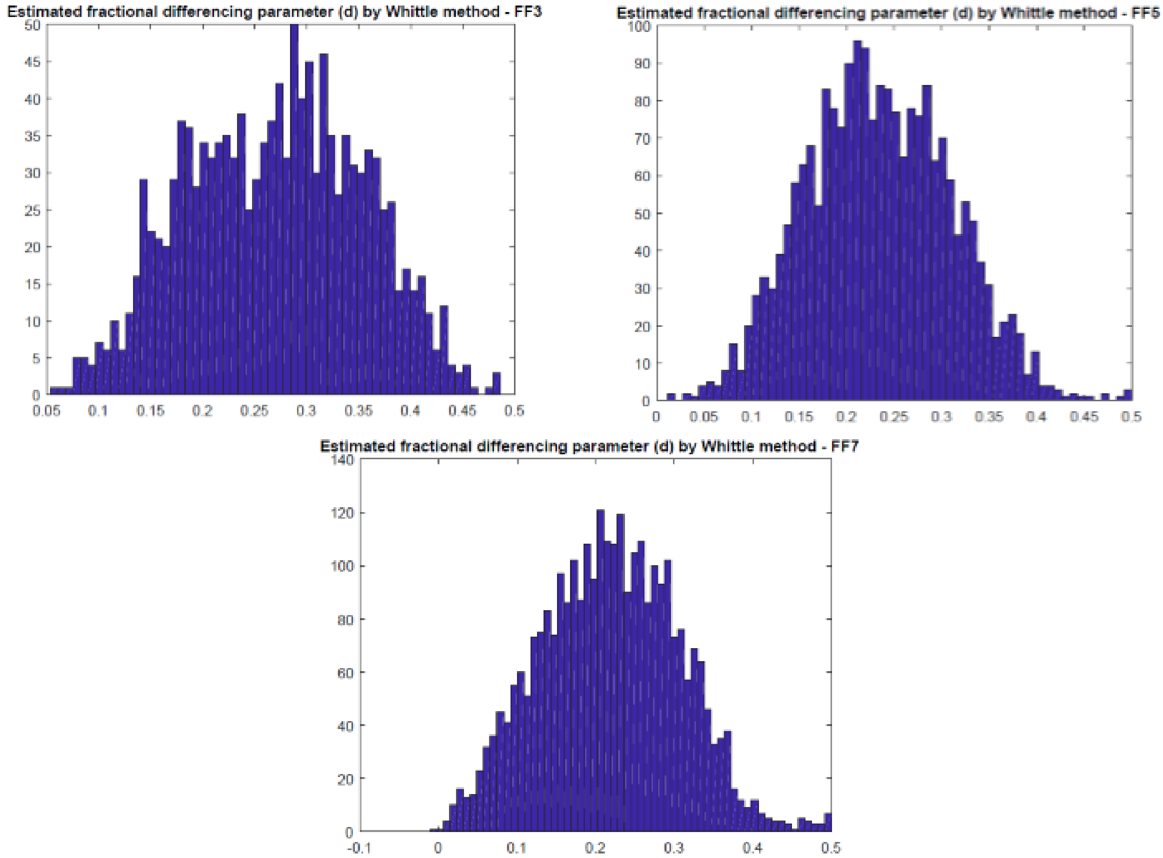
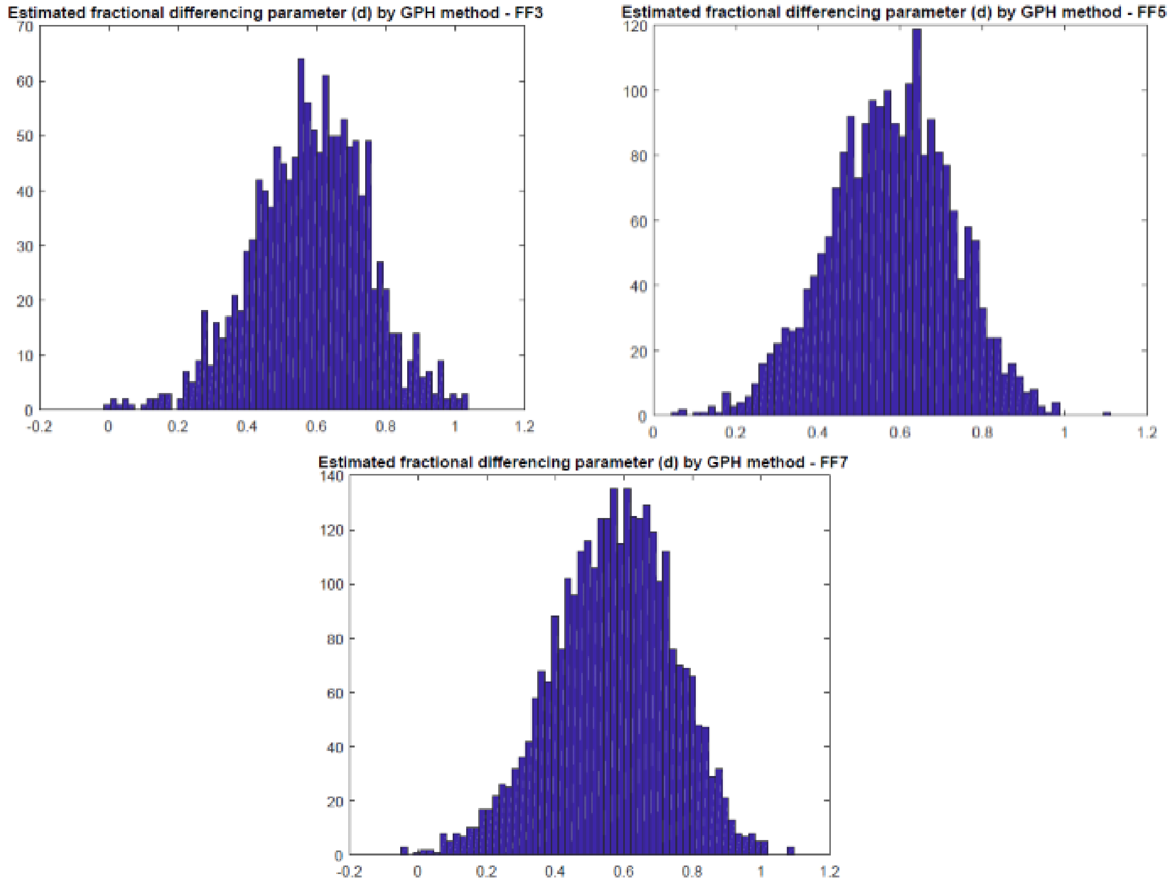


Figure 6: Estimated Fractional Differencing Parameter (d) - GPH Method

These panels show the distribution of fractional differencing parameter (d) estimated by the GPH method. Each panel correspondes to one of the factors configuration (3, 5, or 7 factors). In total, there are $430 \times$ number of factors series of betas for each case. We plot the distribution of d for these series.



C Factors Construction

Market: The market portfolio is a value-weighted portfolio of all stocks in our sample. The market factor is the excess return of this portfolio relative to the risk free rate.

Size and Value: These factors are based on a double-sorting of stocks on market equity and book-to-market equity. Market equity is defined as the the last available price of June times the corresponding value of shares outstanding. Book-to-market ratio is calculated using the book value of the previous fiscal year and the market equity from December of the previous year. Stocks are ranked on market equity and split into two portfolios, small and big (S and B). Book-to-market is used to split the stocks into three book-to-market equity groups based on breakpoints for the bottom 30% (Low, or L), middle 40% (Medium, or M), and high 30% (High, or H). Six portfolios are formed from intersections of the aforementioned groups: S/L, S/M, S/H, B/L, B/M, B/H. Daily value-weighted returns are calculated for the six portfolios, for the whole year (portfolios are rebalanced annually). The factor SMB (small minus big) is the difference between the simple average of returns on the three small-stock portfolios and the simple average on returns of the three big-stock portfolios, that is:

$$SMB = \frac{(R_{S/L} + R_{S/M} + R_{S/H})}{3} - \frac{(R_{B/L} + R_{B/M} + R_{B/H})}{3}$$

Similarly, the factor HML (high minus low) is defined as the difference, daily, of the simple average of the returns on the two high book-to-market portfolios (S/H and B/H) and the simple average of the returns on the two low book-to-market portfolios (S/L and B/L), that is:

$$HML = \frac{(R_{S/H} + R_{B/H})}{2} - \frac{(R_{S/L} + R_{B/L})}{2}$$

Gross Profitability: Follows Novy-Marx (2013). The signal gross profitability (GP) is defined as the ratio between gross profits and total assets. Stocks are ranked and split in 10 deciles every year, using current year financial data. The gross profitability portfolio consists of a strategy that is long in the group with lowest GP and short in the group with the highest GP. For both groups, value-weighted returns are calculated and the factor GP is given by:

$$GP = R_{lowGP} - R_{highGP}$$

Accruals: Follows Sloan (1996). The signal accrual is defined as:

$$\text{Accruals} = \frac{\Delta ACT - \Delta CHE - \Delta LCT + \Delta DLC + \Delta TXP - DP}{(AT + AT_{-12})/2}$$

where ACT is the annual total current assets, CHE is the annual total cash and short-term investments , LCT is the annual current liabilities , DLC is the annual debt in current liabilities , TXP is the annual income taxes payable , DP is the annual depreciation and amortization , and $(AT + AT_{-12})/2$ represents the average total assets over the last two years. Δ stands for annual variation in these variables. The stocks are split in 10 deciles every year, and two value-weighted portfolios are formed on the lowest and highest deciles. The accruals portfolio consists of a strategy that is long on the small accrual portfolio and short on the high accrual portfolio. The factor is then given by:

$$Accrual = R_{lowAccrual} - R_{high_accrual}$$

Asset Growth: Follows Cooper, Gulen, and Schill (2008). The asset growth signal is defined as Asset Growth = AT/AT_{-12}, with AT AT_{-12} previously defined in the accruals section. Analogous to the procedure done for the accruals factor, the asset growth portfolio is simply a strategy that is long on the lowest decile asset growth portfolio and short on the highest decile portfolio (the stocks are sorted into 10 deciles). Following this procedure, the factor is then given by:

$$AssetGrowth = R_{lowAssetGrowth} - R_{highAssetGrowth}$$

Asset Growth: Follows Lyandres, Sun, and Zhang (2008). The investment signal is defined as

$$\text{Investment} = (\Delta PPEGT + \Delta INVT)/AT_{-12}$$

where PPEGT is the gross total property, plant, and equipment (COMPUSTAT's variable 'ppegt'), and INVT is total inventories (COMPUSTAT's 'invt'). For the investment portfolio, the stocks are triple-sorted in size, value (as in the Fama-French factors), and investment. For each of the three characteristics, each stock is classified in one of the three groups: low (30%), medium (30%-70%), or high (70%-100%). This procedure results in 27 different portfolios. The investment portfolio consists of a strategy that is long on the low investment portfolio (the simple average between the 9 groups with low investment) and short on the high investment portfolio (the simple average between the 9 groups with high investment), that is:

$$Investment = R_{lowInvestment} - R_{highInvestment}$$

Risk Free: We get the daily series of returns from Kenneth French website.

D Residual Covariance Matrix - Parameter Selection

In this appendix we show the average equation size and average parameter change for models used to forecast the residual covariance matrices. Since the number of equations is much higher than the one used for the factor covariance matrices, we present the results in percentiles of the distribution (instead of individual plots).

Table 7: Average Equation Size Distribution for Blocks in the Residual Covariance Matrix (3-Factor VHAR)

Distribution	Min	0,25	0,5	0,75	Max	Mean
	Variance Equations					
Group						
Consumer Non-Durables	11.19	11.97	12.19	12.48	13.61	12.29
Consumer Durables	6.25	6.50	6.62	7.00	7.75	6.79
Manufacturing	16.22	16.69	16.85	17.05	17.46	16.86
Oil. Gas. and Coal Extraction	9.00	9.78	10.19	10.59	11.12	10.17
Business Equipment	15.03	15.46	15.64	15.77	16.18	15.62
Telecommunications	6.60	6.90	7.30	7.60	7.90	7.27
Wholesale and Retail	12.80	13.36	13.53	13.71	14.22	13.55
Healthcare. Medical Equipments. and Drugs	11.54	12.19	12.42	12.92	13.58	12.53
Utilities	12.03	12.75	13.00	13.25	14.25	13.03
Others	15.41	15.85	16.47	17.13	18.11	16.53
	Covariance Equations					
Group						
Consumer Non-Durables	0.66	0.83	1.28	1.39	1.52	1.15
Consumer Durables	0.21	0.36	0.54	0.68	0.89	0.53
Manufacturing	0.47	0.55	0.63	0.68	0.72	0.61
Oil. Gas. and Coal Extraction	1.46	1.76	2.09	2.23	2.64	2.01
Business Equipment	0.34	0.40	0.55	0.67	0.71	0.54
Telecommunications	1.36	1.53	1.62	1.73	2.11	1.65
Wholesale and Retail	0.51	0.60	0.89	1.02	1.09	0.83
Healthcare. Medical Equipments. and Drugs	0.42	0.73	1.12	1.58	1.77	1.13
Utilities	0.67	0.74	0.85	1.58	2.06	1.14
Others	0.52	0.60	0.65	0.70	0.72	0.64

Notes: Distribution represents the percentiles and mean value calculated over 473 estimated models. The 430 stocks are separeted in 10 groups, according to SIC. Averages are calculated among stocks in the same group.

Table 8: Average Parameter Change Distribution for Blocks in the Residual Covariance Matrix (3-Factor VHAR)

Distribution	Min	0,25	0,5	0,75	Max	Mean
	Variance Equations					
Group						
Consumer Non-Durables	0.21	0.94	1.25	1.66	3.85	1.33
Consumer Durables	0.00	0.00	0.00	1.56	3.12	0.52
Manufacturing	0.33	0.83	1.02	1.23	3.36	1.08
Oil. Gas. and Coal Extraction	0.00	0.68	1.17	1.56	4.49	1.19
Business Equipment	0.40	0.94	1.18	1.40	3.60	1.21
Telecommunications	0.00	0.00	0.00	1.00	5.00	0.57
Wholesale and Retail	0.25	0.94	1.19	1.43	3.80	1.24
Healthcare. Medical Equipments. and Drugs	0.00	0.74	1.04	1.48	4.88	1.16
Utilities	0.08	0.62	0.93	1.23	4.55	0.98
Others	0.31	0.67	0.80	0.94	1.63	0.82
	Covariance Equations					
Group						
Consumer Non-Durables	0.10	0.25	0.33	0.40	0.64	0.33
Consumer Durables	0.00	0.00	0.45	0.89	3.12	0.50
Manufacturing	0.06	0.09	0.11	0.12	0.20	0.11
Oil. Gas. and Coal Extraction	0.15	0.45	0.56	0.69	1.16	0.58
Business Equipment	0.04	0.08	0.10	0.12	0.20	0.10
Telecommunications	0.00	0.44	0.67	1.11	3.33	0.78
Wholesale and Retail	0.07	0.15	0.18	0.22	0.42	0.19
Healthcare. Medical Equipments. and Drugs	0.05	0.25	0.34	0.45	0.86	0.36
Utilities	0.08	0.21	0.31	0.42	0.78	0.33
Others	0.05	0.06	0.07	0.07	0.10	0.07

Notes: Distribution represents the percentiles and mean value calculated over 473 estimated models. The 430 stocks are separeted in 10 groups, according to SIC. Averages are calculated among stocks in the same group.

Table 9: Average Equation Size Distribution for Blocks in the Residual Covariance Matrix (5 Factors VHAR)

Distribution	Min	0,25	0,5	0,75	Max	Mean
	Variance Equations					
Group						
Consumer Non-Durables	11.32	11.97	12.19	12.45	13.61	12.27
Consumer Durables	6.12	6.50	6.62	7.00	7.62	6.77
Manufacturing	15.91	16.46	16.72	16.97	17.57	16.72
Oil. Gas. and Coal Extraction	10.28	10.75	11.12	11.38	12.06	11.08
Business Equipment	15.00	15.34	15.52	15.72	16.39	15.55
Telecommunications	6.50	6.90	7.30	7.70	7.90	7.30
Wholesale and Retail	12.56	13.22	13.40	13.62	14.24	13.42
Healthcare. Medical Equipments. and Drugs	11.69	12.15	12.54	12.96	13.50	12.55
Utilities	13.47	14.22	14.36	14.56	14.97	14.37
Others	15.35	15.94	16.60	17.22	18.12	16.61
	Covariance Equations					
Group						
Consumer Non-Durables	0.59	0.83	1.33	1.43	1.52	1.17
Consumer Durables	0.21	0.39	0.46	0.57	0.79	0.48
Manufacturing	0.46	0.53	0.61	0.63	0.69	0.59
Oil. Gas. and Coal Extraction	1.38	1.67	1.73	1.95	2.33	1.80
Business Equipment	0.37	0.42	0.57	0.68	0.74	0.56
Telecommunications	1.13	1.44	1.56	1.69	1.91	1.55
Wholesale and Retail	0.59	0.70	0.97	1.06	1.13	0.90
Healthcare. Medical Equipments. and Drugs	0.52	0.75	1.20	1.54	1.78	1.16
Utilities	0.64	0.74	0.95	1.15	1.39	0.95
Others	0.60	0.67	0.70	0.74	0.77	0.70

Notes: Distribution represents the percentiles and mean value calculated over 473 estimated models. The 430 stocks are separeted in 10 groups, according to SIC. Averages are calculated among stocks in the same group.

Table 10: Average Parameter Change Distribution for Blocks in the Residual Covariance Matrix (5-Factor VHAR)

Distribution	Min	0,25	0,5	0,75	Max	Mean
	Variance Equations					
Group						
Consumer Non-Durables	0.10	0.83	1.25	1.56	3.95	1.27
Consumer Durables	0.00	0.00	0.00	0.00	3.12	0.41
Manufacturing	0.31	0.88	1.07	1.30	3.98	1.12
Oil. Gas. and Coal Extraction	0.10	0.59	0.88	1.27	4.49	0.97
Business Equipment	0.27	0.97	1.16	1.37	3.55	1.19
Telecommunications	0.00	0.00	0.00	1.00	6.00	0.61
Wholesale and Retail	0.20	0.94	1.23	1.53	4.10	1.27
Healthcare. Medical Equipments. and Drugs	0.00	0.59	1.04	1.63	4.44	1.17
Utilities	0.00	0.85	1.16	1.47	5.71	1.19
Others	0.42	0.66	0.77	0.91	1.66	0.81
	Covariance Equations					
Group						
Consumer Non-Durables	0.10	0.26	0.33	0.41	0.68	0.33
Consumer Durables	0.00	0.00	0.00	0.45	3.57	0.41
Manufacturing	0.05	0.09	0.10	0.12	0.17	0.10
Oil. Gas. and Coal Extraction	0.14	0.42	0.49	0.58	0.95	0.50
Business Equipment	0.04	0.08	0.10	0.12	0.18	0.10
Telecommunications	0.00	0.22	0.67	0.89	3.11	0.68
Wholesale and Retail	0.08	0.16	0.20	0.24	0.39	0.20
Healthcare. Medical Equipments. and Drugs	0.07	0.26	0.39	0.50	1.11	0.39
Utilities	0.08	0.20	0.28	0.36	0.59	0.28
Others	0.05	0.07	0.07	0.08	0.10	0.07

Notes: Distribution represents the percentiles and mean value calculated over 473 estimated models. The 430 stocks are separeted in 10 groups, according to SIC. Averages are calculated among stocks in the same group.

Table 11: Average Equation Size Distribution for Blocks in the Residual Covariance Matrix (7-Factor HVAR)

Distribution	Min	0,25	0,5	0,75	Max	Mean
	Variance Equations					
Group						
Consumer Non-Durables	11.29	11.94	12.13	12.48	13.74	12.24
Consumer Durables	6.12	6.38	6.50	6.88	7.62	6.65
Manufacturing	15.82	16.38	16.65	16.91	17.37	16.65
Oil. Gas. and Coal Extraction	10.12	10.81	11.03	11.34	11.97	11.09
Business Equipment	14.72	15.20	15.38	15.56	16.05	15.37
Telecommunications	6.40	6.80	7.30	7.50	8.00	7.23
Wholesale and Retail	12.73	13.31	13.47	13.69	14.33	13.49
Healthcare. Medical Equipments. and Drugs	11.77	12.31	12.69	13.15	13.73	12.73
Utilities	13.69	14.22	14.36	14.53	14.92	14.35
Others	15.36	15.93	16.57	17.12	18.05	16.59
	Covariance Equations					
Group						
Consumer Non-Durables	0.59	0.82	1.33	1.41	1.51	1.15
Consumer Durables	0.14	0.36	0.46	0.54	0.75	0.45
Manufacturing	0.50	0.56	0.63	0.65	0.72	0.61
Oil. Gas. and Coal Extraction	1.40	1.64	1.77	2.05	2.42	1.84
Business Equipment	0.54	0.62	0.77	0.90	0.95	0.76
Telecommunications	1.58	1.82	1.91	2.02	2.56	1.93
Wholesale and Retail	0.62	0.73	1.03	1.11	1.16	0.93
Healthcare. Medical Equipments. and Drugs	0.54	0.77	1.02	1.41	1.65	1.06
Utilities	0.78	0.88	1.06	1.12	1.25	1.00
Others	0.63	0.71	0.75	0.78	0.81	0.74

Notes: Distribution represents the percentiles and mean value calculated over 473 estimated models. The 430 stocks are separeted in 10 groups, according to SIC. Averages are calculated among stocks in the same group.

Table 12: Average Parameter Change Distribution for Blocks in the Residual Covariance Matrix (7-Factor HVAR)

Distribution	Min	0,25	0,5	0,75	Max	Mean
	Variance Equations					
Group						
Consumer Non-Durables	0,00	0,94	1,25	1,66	3,64	1,32
Consumer Durables	0,00	0,00	0,00	0,00	3,12	0,40
Manufacturing	0,38	0,88	1,04	1,25	3,74	1,10
Oil, Gas, and Coal Extraction	0,00	0,59	0,88	1,17	3,71	0,92
Business Equipment	0,46	0,91	1,13	1,35	3,14	1,16
Telecommunications	0,00	0,00	0,00	1,00	8,00	0,67
Wholesale and Retail	0,25	0,89	1,19	1,53	3,90	1,24
Healthcare, Medical Equipments, and Drugs	0,00	0,59	1,04	1,63	4,44	1,19
Utilities	0,23	0,85	1,16	1,47	6,10	1,22
Others	0,36	0,65	0,77	0,92	1,73	0,80
	Covariance Equations					
Group						
Consumer Non-Durables	0,05	0,25	0,33	0,40	0,65	0,33
Consumer Durables	0,00	0,00	0,00	0,89	2,23	0,43
Manufacturing	0,06	0,09	0,10	0,12	0,18	0,11
Oil, Gas, and Coal Extraction	0,17	0,43	0,51	0,61	0,92	0,52
Business Equipment	0,06	0,11	0,12	0,14	0,25	0,13
Telecommunications	0,00	0,44	0,89	1,33	4,00	0,93
Wholesale and Retail	0,08	0,17	0,20	0,24	0,40	0,20
Healthcare, Medical Equipments, and Drugs	0,06	0,25	0,34	0,44	0,75	0,35
Utilities	0,10	0,22	0,29	0,36	0,63	0,30
Others	0,05	0,07	0,08	0,08	0,10	0,08

Notes: Distribution represents the percentiles and mean value calculated over 473 estimated models. The 430 stocks are separeted in 10 groups, according to SIC. Averages are calculated among stocks in the same group.

References

Ait-Sahalia, Yacine and Dacheng Xiu (forthcoming). "Using principal component analysis to estimate a high dimensional factor model with high-frequency data". In: *Journal of Econometrics*.

Bauer, G. and K. Vornik (2011). "Forecasting multivariate realized stock market volatility". In: *Journal of Econometrics* 160, pp. 93–101.

Callot, Laurent AF, Anders B Kock, and Marcelo C Medeiros (2017). "Modeling and Forecasting Large Realized Covariance Matrices and Portfolio Choice". In: *Journal of Applied Econometrics* 32.1, pp. 140–158.

Chamberlain, Gary and Michael Rothschild (1983). "Arbitrage, factor structure, and mean-variance analysis on large asset markets". In: *Econometrica* 51.5, p. 1281.

Chiriac, R. and V. Voev (2011). "Modelling and forecasting multivariate realized volatility". In: *Journal of Applied Econometrics* 26, 922–947.

Chiu, Tom YM, Tom Leonard, and Kam-Wah Tsui (1996). "The matrix-logarithmic covariance model". In: *Journal of the American Statistical Association* 91.433, pp. 198–210.

Cooper, Michael J, Huseyin Gulen, and Michael J Schill (2008). "Asset growth and the cross-section of stock returns". In: *The Journal of Finance* 63.4, pp. 1609–1651.

Corsi, Fulvio (2009). "A simple approximate long-memory model of realized volatility". In: *Journal of Financial Econometrics* 7.2, pp. 174–196.

Engle, Robert (2002). "Dynamic Conditional Correlation". In: *Journal of Business & Economic Statistics* 20.3, pp. 339–350.

Engle, Robert and F Kroner (1995). "Multivariate Simultaneous Generalized ARCH". In: *Econometric Theory* 11, pp. 122–150.

Engle, Robert F, Olivier Ledoit, and Michael Wolf (2017). "Large dynamic covariance matrices". In: *Journal of Business & Economic Statistics* forthcoming.

Fama, Eugene F and Kenneth R French (1993). "Common risk factors in the returns on stocks and bonds". In: *Journal of Financial Economics* 33.1, pp. 3–56.

Fan, Jianqing, Yingying Fan, and Jinchi Lv (2008). "High dimensional covariance matrix estimation using a factor model". In: *Journal of Econometrics* 147.1, pp. 186–197.

Fan, Jianquing, Alex Furger, and Dacheng Xiu (20016). "Incorporating Global Industrial Classification Standard into Portfolio Allocation: A Simple Factor-Based Large Covariance Matrix Estimator with High Frequency Data"". In: *Journal of Business & Economic Statistics* 34, pp. 489–503.

Fleming, Jeff, Chris Kirby, and Barbara Ostdiek (2003). "The economic value of volatility timing using "realized" volatility". In: *Journal of Financial Economics* 67.3, pp. 473–509.

Golosnoy, Vasyl, Bastian Gribisch, and Roman Liesenfeld (2012). "The conditional autoregressive Wishart model for multivariate stock market volatility". In: *Journal of Econometrics* 167.1, pp. 211–223.

Lunde, Asger, Neil Shephard, and Kevin Sheppard (2016). "Econometric analysis of vast covariance matrices using composite realized kernels and their application to portfolio choice". In: *Journal of Business & Economic Statistics* 34.4, pp. 504–518.

Lyandres, Evgeny, Le Sun, and Lu Zhang (2008). "The New Issues Puzzle: Testing the Investment-Based Explanation". In: *The Review of Financial Studies* 21.6.

Novy-Marx, Robert (2013). "The other side of value: The gross profitability premium". In: *Journal of Financial Economics* 108.1, pp. 1–28.

Sloan, RG (1996). "Do stock prices fully reflect information in accruals and cash flows about future earnings?" In: *The Accounting Review* 71, pp. 289–315.

Tibshirani, Robert (1996). "Regression shrinkage and selection via the lasso". In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288.

Zou, Hui (2006). "The adaptive lasso and its oracle properies". In: *Journal of the American Statistical Association* 101, pp. 1418–1429.