Journal of Empirical Finance 72 (2023) 23–35

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

Using, taming or avoiding the factor zoo? A double-shrinkage estimator for covariance matrices[☆]

Gianluca De Nard[a,1], Zhao Zhao[b,2,*]

[a] Department of Economics, University of Zurich, CH-8032 Zurich, Switzerland
[b] School of Economics, Huazhong University of Science and Technology, Wuhan, Hubei, China

ARTICLE INFO

Keywords:
Double-shrinkage
Factor models
Markowitz portfolio selection
Multivariate GARCH
Nonlinear shrinkage
Regularization

ABSTRACT

Existing factor models struggle to model the covariance matrix for a large number of stocks and factors. Therefore, we introduce a new covariance matrix estimator that first shrinks the factor model coefficients and then applies nonlinear shrinkage to the residuals and factors. The estimator blends a regularized factor structure with conditional heteroskedasticity of residuals and factors and displays superior all-around performance against various competitors. We show that for the proposed double-shrinkage estimator, it is enough to use only the market factor or the most important latent factor(s). Thus there is no need for laboriously taking into account the factor zoo.

1. Introduction

Factor models have a long history in finance, with a wide range of applications both in theory (e.g., Sharpe, 1963; Ross, 1976; Fama and French, 1992) and in practice (e.g., Meucci, 2005; Chincarini and Kim, 2006). On the one hand, the search for observed factors that explain the cross-section of expected stock returns has produced hundreds of potential candidates, as noted by Cochrane (2011) and more recently by Harvey and Liu (2015), McLean and Pontiff (2016), and Hou et al. (2020). On the other hand, common latent factors are typically extracted from the covariance matrix of returns (see, e.g., Kelly et al., 2019; Gu et al., 2021) without any need of external data. In contrast to the booming factor literature on expected stock returns, much less has been done on the estimation of covariance matrix of stock returns, which plays an important role in portfolio selection and risk management. In this article, we analyze the usefulness of the numerous observed factors and compare it with latent factor models based only on the underlying return data for the estimation of large-dimensional covariance matrices.

In the framework of dynamic factor models, we propose a double-shrinkage estimator, which first shrinks the coefficients based on penalized linear factor models (ridge, LASSO and elastic nets) and afterwards uses the nonlinear shrinkage method to estimate the covariance matrices of both, factors and residuals. The purposes of the first and the second shrinkage methods are to address the curse of dimensionality from the large number of factors and stocks, respectively.

We run an extensive back-test on historical data based on a large data set of U.S. stock returns and 99 common factors dating back 40 years. Empirical results show that the positive effects of both shrinkage steps on the out-of-sample performance are statistically and economically significant. Moreover, approximate factor models always beat exact factor models, regardless of the number of

☆ We thank Michael Wolf, Olivier Ledoit, Bryan Kelly, Sebastian Stöckl, Nikolaus Hautsch, and participants at the 13th Annual SoFiE Conference. We gratefully acknowledge the National Natural Science Foundation of China (Grant No. 72173048), the Fundamental Research Funds for the Central Universities (Grant No. 2021XXJS054) and the Swiss National Science Foundation.

* Corresponding author.
E-mail addresses: gianluca.denard@econ.uzh.ch, denard@stern.nyu.edu (G. De Nard), zhaozhao@hust.edu.cn (Z. Zhao).
1 Further affiliations: New York University, NYU Stern Volatility and Risk Institute and OLZ AG.
2 Second affiliation: Research Center for Modern Economics.

https://doi.org/10.1016/j.jempfin.2023.02.003
Received 14 May 2022; Received in revised form 19 October 2022; Accepted 19 February 2023
Available online 22 February 2023
0927-5398/

factors. We find that approximate factor models do not profit from a large number of factors, and it is enough to use the Fama–French factors, or even only the market factor, in our approximate factor model based on the double-shrinkage estimation: ridge regularization in conjunction with DCC-NL of Engle et al. (2019). Additionally, the proposed double-shrinkage estimator works even better for latent factor models. To sum up, taming the factor zoo is clearly beneficial, however, avoiding it is even better.[3] Consequently, we suggest focusing only on the information in the underlying return time series and simply computing the (optimal number of) latent factors. This finding makes the double-shrinkage estimator even more attractive for industry and applied portfolio management, since only return data are needed.

A further contribution of this paper is to propose a new forecasting scheme in the case where a dynamic covariance matrix estimator is used and the holding period of the portfolio exceeds the frequency of the observed returns; for example, when the frequency of the observed returns is daily but the portfolio is held for a month to reduce turnover. In such a case, De Nard et al. (2021) recommend an "averaged forecast" of the covariance matrix by averaging over covariance matrix forecasts. We propose an improved estimation scheme by averaging forecasted portfolio weights instead. Unlike the dynamic weighting schemes for characteristic-based portfolios proposed by Lioui and Tarelli (2020), our scheme focuses on the dynamics of factors and residuals rather than the market exposures, as we have a purer goal of estimating the covariance matrix and a shorter investment horizon, where the fluctuation of market exposures is less important.

The literature on Markowitz portfolio selection in large dimensions has experienced a large expansion over the last decade or so and is still growing.[4] Typically, the idea is to constrain the portfolio weights directly or indirectly through shrinking the covariance matrix to reduce estimation error and potential extreme positions. For example, Ledoit and Wolf (2017) use a nonlinear shrinkage estimator of the covariance matrix to compute the global minimum variance portfolio and show that it performs better than previous state-of-the-art estimators. More recently, De Nard et al. (2021) combine factor model with DCC-NL, and conclude that the new estimator delivers more efficient portfolio selection. Our new approach inherits the idea of shrinkage, and goes further by regularizing numerous risk factors.

Researchers have found a benefit of using observed and latent factor models relative to structure-free models (i.e., the intercept-only model in our definition) in the estimation of covariance matrices; for example, see Fan et al. (2008, 2013, 2016) and the references therein. In recent years, the application and the corresponding statistical theory have been extended to high-frequency factor models; for example, see Fan and Kim (2018, 2019), Kim and Fan (2019), and Sun and Xu (2021). Another recent development is the consideration of time-varying conditional heteroscedasticity, normally using a MGARCH model (e.g., So et al., 2020; De Nard et al., 2021). However, the literature so far only considers one or several market indices as observed factor(s) and does not tell us how well factor models would do if a large set of risk factors are incorporated. As there is a little consensus in the literature about the nature and the number of such observed risk factors to be used, except for the market and Fama–French factors, we investigate the factors beyond the five-factor model of Fama and French (2015) by taking into account the entire "factor zoo" of Feng et al. (2020). Therefore, in contrast to prior studies, in this paper *large dimension* refers not only to a large number of stocks, but also to a large number of observed factors.[5]

The remainder of the paper is organized as follows. Section 2 gives a description of regularized factor models. Section 3 details our new dynamic estimation schemes for estimating large-dimensional covariance matrices with factor models. Section 4 describes the empirical methodology and presents the results of the out-of-sample backtest exercise based on real-life stock return data. Section 5 concludes. Supplementary material provides additional details and results.

2. Factor models

As stated before, we are interested in estimating covariance matrices when the number of assets, N, is of the same order of magnitude as the number of observations, T, for a large number of potential factors, K.

2.1. Notation

In what follows, the subscript i indexes the assets and covers the range of integers from 1 to N, where N denotes the dimension of the investment universe; the subscript k indexes the factors and covers the range of integers from 1 to K, where K denotes the number of factors; the subscript t indexes the dates and covers the range of integers from 1 to T, where T denotes the sample size. The notation $\mathsf{Cor}(\cdot)$ represents the correlation matrix of a random vector, the notation $\mathsf{Cov}(\cdot)$ represents the covariance matrix of a random vector, and the notation $\mathsf{Diag}(\cdot)$ represents the function that sets to zero all the off-diagonal elements of a matrix. Furthermore, we use the following notations:

- $r_{i,t}$: return for asset i at date t, stacked into $r_t := (r_{1,t}, \dots, r_{N,t})'$
- $f_{k,t}$: factor k at date t, stacked into $f_t := (f_{1,t}, \dots, f_{K,t})'$

[3] By "avoiding", we do not suggest using a structure-free estimator without any factor, e.g. the DCC-NL of Engle et al. (2019) but recommend focusing only on the market factor, or even better, use latent factors. We argue that there is no need for laboriously computing all the published accounting and finance factors.

[4] To list only a few examples, see Ledoit and Wolf (2003), Jagannathan and Ma (2003), Fan et al. (2008), DeMiguel et al. (2009a), Fan et al. (2013), DeMiguel et al. (2013), Ledoit and Wolf (2017), Bollerslev et al. (2018), Han (2020), (De Nard et al., 2021), Han et al. (2021), De Nard et al. (2022) and De Nard (2022).

[5] Note that in a different context factors are also used to explain first moments, that is, to explain the cross-section of expected stock returns; for example, see Feng et al. (2020). However, factors that are successful in this context are not necessarily useful for explaining second moments.

- $u_{i,t}$: error term for asset i at date t, stacked into $u_t := (u_{1,t}, \dots, u_{N,t})'$
- $x_{j,t}$: underlying time series for covariance matrix estimation, thus $x_{j,t} \in \{f_{j,t}, u_{j,t}\}$, stacked into $x_t := (x_{1,t}, \dots, x_{J,t})'$; where $J = N$ for u_t; and $J = K$ for f_t
- $d^2_{j,t} := \mathrm{Var}(x_{j,t}|\mathcal{F}_{t-1})$: conditional variance of the jth variable at t
- $s_{j,t} := x_{j,t}/d_{j,t}$: devolatized series, stacked into $s_t := (s_{1,t}, \dots, s_{J,t})'$
- $w_{i,t}$: portfolio weight for asset i at date t, stacked into $w_t := (w_{1,t}, \dots, w_{N,t})'$
- D_t: the J-dimensional diagonal matrix whose jth diagonal element is $d_{j,t}$
- $R_t := \mathrm{Cor}(x_t|\mathcal{F}_{t-1}) = \mathrm{Cov}(s_t|\mathcal{F}_{t-1})$: conditional correlation matrix at date t
- $\Sigma_{x,t} := \mathrm{Cov}(x_t|\mathcal{F}_{t-1})$: conditional covariance matrix at date t; thus $\mathrm{Diag}(\Sigma_t) = D^2_t$
- $C := \mathbb{E}(R_t) = \mathrm{Cor}(x_t) = \mathrm{Cov}(s_t)$: unconditional correlation matrix

2.2. Factor zoo

In our description below, the factors are allowed to be observed, such as Fama–French factors, or latent, which are estimated from historical return data in a prior step. There is an important distinction between observed and latent factors. Latent factors are unknown and must be estimated from the data along with the factor loadings. The two most popular methods to this end are maximum likelihood estimation (MLE) and principal components analysis (PCA); see Bai and Shi (2011, Section 5). In our empirical analysis, we use the method of PCA, which leads to consistent estimation of the factors (up to a rotation) under suitable regularity conditions; see Fan et al. (2013, Section 2). We tacitly assume such a set of conditions.

Observed factors are known and based on outside information. The leading example is the one of Fama–French factors. Many other factors have been proposed in the literature; for example, see Bai and Shi (2011, Section 4) and Feng et al. (2020). In contrast to De Nard et al. (2021), we are interested in large factor models beyond the Fama–French factors. More specifically, we investigate if a (regularized) factor model can benefit from the big data information of the factor zoo, or if the signal-to-noise ratio is too weak for estimating covariance matrices. To this end, we add the 94 factors of Gu et al. (2020) to the five Fama–French factors, resulting in a large factor universe of $K = 99$.

Note that for the factor models of interest, discussed in the next sections, we need common factors instead of stock-level characteristics (factor scores) as used and published by Gu et al. (2020). Therefore, we first need to compute common factors out of the stock-level characteristics. Going back to at least Fama and French (1993), the preferred method for computing common factors has been to construct portfolios based on sorting. The anomaly-replication literature focuses on dollar-neutral long-short portfolios by going long the stocks that are in the top quintile according to their stock-level characteristics, and short the stocks in the bottom quintile.[6] Additionally, the literature suggests using NYSE breakpoints and value-weighted returns to form the long-short portfolio to control for microcaps biases; see Cremers et al. (2012), Fama and French (2008), Feng et al. (2020), Freyberger et al. (2020) and Hou et al. (2020) among others.

On the other hand, Ledoit et al. (2019) propose a different approach to replicate common factors. They argue that the traditional finance literature on replicating anomalies ignores any information on the covariance matrix of stock returns and is therefore suboptimal. Historically, it has been difficult to estimate the covariance matrix for a large universe of stocks. However, they demonstrate that using the recent DCC-NL estimator of the covariance matrix of Engle et al. (2019) in a large investment universe multiplies the "Student" t-statistics for cross-sectional anomaly detection, on average, by a factor of more than two relative to the status quo. Therefore, it is in everybody's interest to upgrade the theoretically and empirically underpowered portfolio-construction procedure based on sorting. However, their efficient sorting method is based on equally-weighted returns and does not consider NYSE breakpoints. Therefore, we combine the traditional quantile based long-short portfolio, controlling for microcap biases via NYSE breakpoints and value-weighted returns, with the efficient sorting method of Ledoit et al. (2019), taking into account the information of the covariance matrix of the stock returns. Supplementary material provides a detailed description on how to compute the 94 factors of Gu et al. (2020) based on efficient sorting.

2.3. Regularized factor models

Assume that for every asset $i = 1, \dots, N$,

$$r_{i,t} = \alpha_i + \beta_i' f_t + u_{i,t} \ , \tag{2.1}$$

with $\beta_i := (\beta_{i,1}, \dots, \beta_{i,K})'$ and $\mathbb{E}(u_{i,t}|f_t) = 0$. Furthermore, letting $u_t := (u_{1,t}, \dots, u_{N,t})'$. The covariance matrices of f_t and u_t may both be time-varying. The time-varying (conditional) covariance matrix of r_t is given by

$$\Sigma_{r,t} = B' \Sigma_{f,t} B + \Sigma_{u,t} \ , \tag{2.2}$$

where B is the $K \times N$ matrix whose ith column is the vector β_i. The estimator of the time-varying conditional covariance matrix of r_t is then given by

$$\hat{\Sigma}_{r,t} := \hat{B}' \hat{\Sigma}_{f,t} \hat{B} + \hat{\Sigma}_{u,t} \ . \tag{2.3}$$

[6] Instead of quintiles, some authors may prefer terciles or deciles.

As we are interested in estimating large-dimensional covariance matrices with the help of a large number of (common) factors, not only the (dynamic) residual covariance matrix, $\Sigma_{u,t}$, but also the (dynamic) factor covariance matrix, $\Sigma_{f,t}$, is large. The simple linear (OLS) model is expected to fail in the presence of many factors. When the number of factors K approaches the number of observations T, the linear model becomes inefficient or even inconsistent. It begins to overfit noise rather than extracting signal. Crucial for avoiding overfit is reducing the number of estimated parameters. The most common machine learning device for imposing parameter parsimony is to append a penalty to the objective function in order to favor more parsimonious specifications. This "regularization" of the estimation problem mechanically deteriorates a model's in sample performance in hopes of improving its stability out of sample. This will be the case when penalization manages to reduce the model's fit of the noise while preserving its fit of the signal.

Penalized linear models use shrinkage and variable selection to manage large dimensionality by forcing the coefficients on most regressors near or exactly to zero. Arguably, ridge and LASSO regressions are the standard machine learning methods to reduce the number of estimated parameters, respectively the curse of dimensionality, and thus avoiding overfitting. Note that a convex combination of the two penalized methods is known as the elastic net.

2.3.1. Ridge factor model

Ridge shrinks the regression coefficients towards zero by imposing a L^2 penalty on their size. The ridge coefficients minimize a penalized residual sum of squares

$$\left[\hat{\alpha}_i^{\text{ridge}}, \hat{\beta}_i^{\text{ridge}}\right] := \underset{\alpha_i,\beta_i}{\operatorname{argmin}} \left\{ \sum_{t=1}^{T} \left(r_{i,t} - \alpha_i - \sum_{k=1}^{K} f_{t,k}\beta_{i,k} \right)^2 + \lambda \sum_{k=1}^{K} \beta_{i,k}^2 \right\}. \tag{2.4}$$

Here, $\lambda \geq 0$ is a penalty parameter that controls the amount of shrinkage: the larger the value of λ, the greater the amount of shrinkage. In the special case $\lambda = 0$, we obtain the basic OLS coefficients; whereas $\lambda = \infty$ returns the (fully shrunk) intercept-only model.

Finally, we collect the time-invariant estimated ridge betas across assets into a $K \times N$ matrix $\hat{B}^{\text{ridge}}$, where $\hat{\beta}_i^{\text{ridge}}$ is the ith column of $\hat{B}^{\text{ridge}}$. Under the model assumptions, the estimated time-varying conditional covariance matrix of r_t is then given by

$$\hat{\Sigma}_{r,t} := \hat{B}^{\text{ridge}\prime}\, \hat{\Sigma}_{f,t} \hat{B}^{\text{ridge}} + \hat{\Sigma}_{u,t}^{\text{ridge}} \ . \tag{2.5}$$

As explained above, the estimation of time-varying covariance matrices for the factors and (vectors of) ridge regression errors are detailed in Section 3.

Remark 2.1 (*Shifted Ridge*). We also consider an alternative formulation of ridge, where we shift its shrinkage target. Thereby, we shrink the coefficients not towards zero, but towards their cross-sectional mean $\bar{\beta}_k := 1/N \sum_{i=1}^{N} \beta_{i,k}$:

$$\left[\hat{\alpha}_i^{\text{Sridge}}, \hat{\beta}_i^{\text{Sridge}}\right] := \underset{\alpha_i,\beta_i}{\operatorname{argmin}} \left\{ \sum_{t=1}^{T} \left(r_{i,t} - \alpha_i - \sum_{k=1}^{K} f_{t,k}\beta_{i,k} \right)^2 + \lambda \sum_{k=1}^{K} (\beta_{i,k} - \bar{\beta}_k)^2 \right\} . \tag{2.6}$$

However, in unreported results we find no significant benefit of shrinking the coefficients towards their cross-sectional mean, at least for the estimation of covariance matrices.

2.3.2. LASSO factor model

The LASSO (least absolute shrinkage and selection operator) is a shrinkage method similar to ridge, with subtle but important differences. LASSO includes a L^1 penalization that makes the solution nonlinear without closed-form expression. LASSO is a variable-selection method that imposes sparsity on the specification and sets coefficients on a subset of covariates *exactly* to zero, where λ controls again the amount of shrinkage:

$$\left[\hat{\alpha}_i^{\text{LASSO}}, \hat{\beta}_i^{\text{LASSO}}\right] := \underset{\alpha_i,\beta_i}{\operatorname{argmin}} \left\{ \sum_{t=1}^{T} \left(r_{i,t} - \alpha_i - \sum_{k=1}^{K} f_{t,k}\beta_{i,k} \right)^2 + \lambda \sum_{k=1}^{K} |\beta_{i,k}| \right\} . \tag{2.7}$$

Finally, we collect the time-invariant estimated LASSO betas across assets into a $K \times N$ matrix $\hat{B}^{\text{LASSO}}$, where $\hat{\beta}_i^{\text{LASSO}}$ is the ith column of $\hat{B}^{\text{LASSO}}$. Under the model assumptions, the estimated time-varying conditional covariance matrix of r_t is then given by

$$\hat{\Sigma}_{r,t} := \hat{B}^{\text{LASSO}\prime}\, \hat{\Sigma}_{f,t} \hat{B}^{\text{LASSO}} + \hat{\Sigma}_{u,t}^{\text{LASSO}} \ . \tag{2.8}$$

As explained above, the estimation of time-varying covariance matrices for the factors and (vectors of) LASSO regression errors are detailed in Section 3.

2.3.3. Elastic net factor model

We also consider a convex combination of the two penalties

$$\left[\hat{\alpha}_i^{\text{ENet}}, \hat{\beta}_i^{\text{ENet}}\right] := \underset{\alpha_i,\beta_i}{\operatorname{argmin}} \left\{ \sum_{t=1}^{T} \left(r_{i,t} - \alpha_i - \sum_{k=1}^{K} f_{t,k}\beta_{i,k} \right)^2 + \lambda \sum_{k=1}^{K} (\gamma \beta_{i,k}^2 + (1-\gamma)|\beta_{i,k}|) \right\} , \tag{2.9}$$

where the elastic net involves two non-negative parameters, λ and γ. The $\gamma = 0$ case corresponds to the LASSO and uses an absolute value parameter penalization, thus it sets coefficients on a subset of factors exactly to zero. In this sense, $\gamma = 0$ imposes sparsity on the specification and can thus be thought of as a variable selection method. The $\gamma = 1$ case corresponds to ridge regression that draws all coefficient estimates closer to zero but does not impose exact zeros anywhere. In this sense, $\gamma = 1$ is a shrinkage method that helps prevent coefficients from becoming unduly large in magnitude. For intermediate values of γ, the elastic net encourages simple models through both shrinkage and selection. We adaptively optimize the tuning parameters, λ and γ, using the validation sample. For a detailed description of the estimation of regularization parameters, see the supplementary material. It describes how we design disjoint sub-samples for the estimation of the introduced regularized factor models and our notion of "parameter tuning".

Finally, we collect the time-invariant estimated elastic net betas across assets into a $K \times N$ matrix $\hat{B}^{\text{ENet}}$, where $\hat{\beta}_i^{\text{ENet}}$ is the ith column of $\hat{B}^{\text{ENet}}$. Under the model assumptions, the estimated time-varying conditional covariance matrix of r_t is then given by

$$\hat{\Sigma}_{r,t} := \hat{B}^{\text{ENet}\prime}\, \hat{\Sigma}_{f,t}\, \hat{B}^{\text{ENet}} + \hat{\Sigma}_{u,t}^{\text{ENet}} \; . \tag{2.10}$$

3. New dynamic estimation schemes

We use daily data to forecast covariance matrices but then hold the portfolio for an entire month before updating it again. This creates a certain "mismatch" for dynamic models, which assume that the (conditional) covariance matrix changes at the forecast frequency, that is, at the daily level: Why use a covariance matrix forecasted only for the next day to construct a portfolio that will then be held for an entire month?

To address this mismatch, De Nard et al. (2021) present an "averaged-forecasting" approach for dynamic models: At portfolio construction date h, forecast the covariance matrix for all days of the upcoming month, that is, for $t = h, h+1, \ldots, h+20$; then average those 21 forecasts and use this "averaged forecasts" to construct the portfolio at date h.

As an alternative, we propose to average portfolio weights instead of covariance matrices. More specifically, at portfolio construction date h, forecast the covariance matrix for all days of the upcoming month, that is, for $t = h, h+1, \ldots, h+20$; then compute the 21 portfolios for each forecasted covariance matrix; finally "average" those 21 portfolios to construct the portfolio at date h.

For the dynamics of the univariate volatilities of the underlying time series $x_t \in \{f_t, u_t\}$, we use a GARCH(1,1) process:[7]

$$d_{j,t}^2 = \omega_j + \delta_{1,j} x_{j,t-1}^2 + \delta_{2,j} d_{j,t-1}^2 \; , \tag{3.1}$$

where $(\omega_j, \delta_{1,j}, \delta_{2,j})$ are the variable-specific GARCH(1,1) parameters. We assume that the evolution of the conditional correlation matrix over time is governed as in the DCC-NL model of Engle et al. (2019)[8]:

$$Q_t = (1 - \delta_1 - \delta_2)C + \delta_1 s_{t-1} s_{t-1}' + \delta_2 Q_{t-1} \; , \tag{3.2}$$

where (δ_1, δ_2) are the DCC-NL parameters analogous to $(\delta_{1,j}, \delta_{2,j})$. The matrix Q_t can be interpreted as a conditional pseudo-correlation matrix, or a conditional covariance matrix of devolatized residuals. It cannot be used directly because its diagonal elements, although close to one, are not exactly equal to one. From this representation, we obtain the conditional correlation matrix and the conditional covariance matrix as

$$R_t := \text{Diag}(Q_t)^{-1/2}\; Q_t\; \text{Diag}(Q_t)^{-1/2} \tag{3.3}$$

$$\Sigma_{x,t} := D_t R_t D_t \; . \tag{3.4}$$

Hence, to determine the portfolio of interest, we average the L forecasted portfolio weights based on the L forecasted conditional covariance matrices $\hat{\Sigma}_{x,h+l} = \hat{D}_{h+l} \hat{R}_{h+l} \hat{D}_{h+l}$, for $l = 0, 1, \ldots, L-1$.[9] For factor models, both, the conditional covariance matrix of the factors, $\Sigma_{f,h+l}$, and the conditional covariance matrix of the residuals, $\Sigma_{u,h+l}$, need to be forecasted first to compute the forecasted return covariance matrix:

$$\hat{\Sigma}_{r,h+l} := \hat{B} \hat{\Sigma}_{f,h+l} \hat{B} + \hat{\Sigma}_{u,h+l} \; . \tag{3.5}$$

As we discuss in the next section, we are interested in a "clean" portfolio in terms of evaluating the quality of a covariance matrix estimator and thus focus mostly on the global minimum variance portfolio

$$\hat{w}_{h+l} := \frac{\hat{\Sigma}_{r,h+l}^{-1} \mathbb{1}}{\mathbb{1}' \hat{\Sigma}_{r,h+l}^{-1} \mathbb{1}} \; , \tag{3.6}$$

see Eq. (4.3). Therefore, to get the frequency adjusted portfolio on portfolio-construction day h, we average over the L forecasted portfolios:

$$\hat{w}_h^* := \frac{1}{L} \sum_{l=0}^{L-1} \hat{w}_{h+l} \; . \tag{3.7}$$

[7] Note that we define the same GARCH(1,1) model for all underlying time series. Thus the variance of the factors, residuals, and returns (for structure-free estimators as discussed in Remark 4.2) are based on the same model parametrization. Of course, their variable-specific GARCH(1,1) parameter estimates differ across underlyings.

[8] Note that the DCC-NL is a DCC with unconditional correlation matrix estimated by the nonlinear shrinkage estimator.

[9] The supplementary material provides a detailed description on how to forecast D_{h+l} and R_{h+l}.

Note that in practice, the GARCH parameters and the DCC(-NL) parameters need to be estimated first. In doing so, we mainly follow the suggestions of Engle et al. (2019, Section 3).

First, the GARCH parameters of Eq. (3.1) are estimated using (pseudo) maximum likelihood assuming normality. This results in estimators $(\hat{\omega}_j, \hat{\delta}_{1,j}, \hat{\delta}_{2,j})$ that are used for devolatizing returns and are also used for forecasting conditional variances.

Second, the correlation-targeting matrix C of Eq. (3.2) is estimated via nonlinear shrinkage applied to the $\{s_t\}$, with post-processing to enforce a proper correlation matrix; to speed up the computations, we use the analytical nonlinear shrinkage method of Ledoit and Wolf (2020).[10] Having an estimator $\hat{C}$, in one of these two ways, we then estimate the DCC parameters (δ_1, δ_2) of Eq. (3.2) using the (pseudo) composite likelihood method of Pakel et al. (2021) assuming normality.[11] In this way, $(\hat{\omega}_j, \hat{\delta}_{1,j}, \hat{\delta}_{2,j}, \hat{\delta}_1, \hat{\delta}_2)$ are used for forecasting conditional correlation matrices. Combining forecasts of conditional variances with forecasts of conditional correlation matrices yields forecasts of conditional covariance matrices.

4. Empirical analysis

4.1. Data and portfolio-construction rules

We restrict attention to stocks from the NYSE, AMEX, and NASDAQ stock exchanges. We download daily stock return data from the Center for Research in Security Prices (CRSP) starting in 01/01/1977 and ending in 12/31/2016. We also download daily returns on the five factors of Fama and French (2015) from the website of Ken French. In addition, we obtain the 94 (monthly stock-level) factor scores used by Gu et al. (2020) from Dacheng Xiu's webpage.[12] The supplementary material lists all the factors including their corresponding main literature and details the methodology to compute the common factors.

For simplicity, we adopt the common convention that 21 consecutive trading days constitute one "month". The out-of-sample period ranges from 01/01/1982 through 12/31/2016, resulting in a total of 420 "months" (or 8820 days). All portfolios are updated monthly.[13] We denote the investment dates by $h = 1, \dots, 420$. At any investment date h, a covariance matrix is estimated based on the most recent 1260 daily returns, which roughly corresponds to using five years of past data.

We consider the following portfolio sizes: $N \in \{100, 500, 1000\}$. For a given combination (h, N), the investment universe is obtained as follows. We find the set of stocks that have an almost complete return history over the most recent $T = 1260$ days as well as a complete return "future" over the next 21 days.[14] We then look for possible pairs of highly correlated stocks, that is, pairs of stocks that have returns with a sample correlation exceeding 0.95 over the past 1260 days. In such pairs, if they should exist, we remove the stock with the lower market capitalization of the two on investment date h. The reason is that we do not want to include highly similar stocks. Of the remaining set of stocks, we then pick the largest N stocks (as measured by their market capitalization on investment date h) as our investment universe. In this way, the investment universe changes relatively slowly from one investment date to the next.

There is a great advantage in having a well-defined rule that does not involve drawing stocks at random, as such a scheme would not have to be replicated many times and averaged over to give stable results. As far as rules go, the one we have chosen seems the most reasonable because it avoids so-called "penny stocks" whose behavior is often erratic; also, high-market-cap stocks tend to have the lowest bid–ask spreads and the highest depth in the order book, which allows large investment funds to invest in them without breaching standard safety guidelines.

We consider the problem of estimating the global minimum variance (GMV) portfolio in the absence of short-sales constraints. The problem is formulated as

$$\min_{w} w' \Sigma_{r,t} w \tag{4.1}$$

$$\text{subject to} \quad w' \mathbb{1} = 1 \,, \tag{4.2}$$

where $\mathbb{1}$ denotes a vector of ones of dimension $N \times 1$. It has the analytical solution

$$w = \frac{\Sigma_{r,t}^{-1} \mathbb{1}}{\mathbb{1}' \Sigma_{r,t}^{-1} \mathbb{1}} \,. \tag{4.3}$$

The natural strategy in practice is to replace the unknown $\Sigma_{r,t}$ by an estimator $\hat{\Sigma}_{r,t}$ in formula (4.3), yielding a feasible portfolio. As all portfolios are updated on a monthly basis, but the conditional covariance matrices change at the daily level, we follow the new estimation scheme by averaging over all daily forecasted portfolios of the particular month:

$$\hat{w} := \frac{1}{L} \sum_{l=0}^{L-1} \frac{\hat{\Sigma}_{r,t+l}^{-1} \mathbb{1}}{\mathbb{1}' \hat{\Sigma}_{r,t+l}^{-1} \mathbb{1}} \,. \tag{4.4}$$

[10] In contrast, Engle et al. (2019, Section 3) used the numerical method of Ledoit and Wolf (2015).

[11] As Engle et al. (2019, Section 3) do, we using neighboring pairs of assets to build up a (pseudo) composite likelihood.

[12] See, http://dachxiu.chicagobooth.edu.

[13] Monthly updating is common practice to avoid an unreasonable amount of turnover and thus transaction costs. During a month, from one day to the next, we hold number of shares fixed rather than portfolio weights; in this way, there are no transactions at all during a month.

[14] The first restriction allows for up to 2.5% of missing returns over the most recent 1260 days, and replaces missing values by zero. The latter, "forward-looking" restriction is not a feasible one in real life but is commonly applied in the related finance literature on the out-of-sample evaluation of portfolios.

Estimating the GMV portfolio is a "clean" problem in terms of evaluating the quality of a covariance matrix estimator, since it abstracts from having to estimate the vector of expected returns at the same time. In addition, researchers have established that estimated GMV portfolios have desirable out-of-sample properties not only in terms of risk but also in terms of reward-to-risk, that is, in terms of the information ratio; for example, see Haugen and Baker (1991), Jagannathan and Ma (2003), and Nielsen and Aylursubramanian (2008). As a result, such portfolios have become additions to the large array of products sold by the mutual-fund industry. In addition to Markowitz portfolios based on formula (4.4), we also include as a simple-minded benchmark the equal-weighted portfolio promoted by DeMiguel et al. (2009b), among others, since it has been claimed to be difficult to outperform. We denote the equal-weighted portfolio by $\mathbf{1/N}$.

4.2. Competing covariance matrix estimators

We now detail the various covariance matrix estimators included in our empirical analysis. We focus on the regularization method (OLS, ridge, LASSO or elastic net), on the size of the factor model ($K \in \{1, 3, 5, 48, 99\}$), on the structure of the factor model (exact or approximate factor model), and on the factor type (observed vs. latent).

For the comparison of small ($K = 1$) and large ($K = 99$) factor models, we use the DCC-NL covariance matrix estimator of Engle et al. (2019) for $\hat{\Sigma}_{f,t}$, as it performs well across all dimensions.[15] More specifically, $K = 1$ is the one-factor model based on the first Fama–French factor; $K = 3$ is the Fama–French three-factor model; $K = 5$ is the Fama–French five-factor model; $K = 99$ is the large factor model including all five Fama–French factors and the 94 analyzed factors. We also report the large $K = 48$ factor model based on the five Fama–French factors and the 43 highly significant factors detailed in the supplementary material.[16]

The main difference between an exact and approximate factor model lies in the estimation of the residual covariance matrix $\Sigma_{u,t}$. For an exact factor model the diagonal sample covariance matrix $\mathsf{Diag}(S_{\hat{u}})$ is used, without any shrinkage or GARCH model, and for an approximate factor model we apply the DCC-NL estimator of Engle et al. (2019) to $\{\hat{u}_t\}$.[17]

Unless stated otherwise, the estimators are based on the new (portfolio averaging) estimation scheme of Section 3. The first four estimators in our following list are from exact factor models, whereas the remaining ones are based on approximate factor models.

- **EFM:** estimator based on an exact (OLS) factor model as in Formula (2.3).
- **ERFM:** estimator based on an exact (**R**IDGE) factor model as in Formula (2.5).
- **ELFM:** an estimator based on an exact (**L**ASSO) factor model as in Formula (2.8).
- **ENFM:** estimator based on an exact (Elastic **N**et) factor model as in Formula (2.10).
- **AFM:** estimator based on an approximate (OLS) factor model as in Formula (2.3).
- **ARFM:** estimator based on an approximate (**R**IDGE) factor model as in Formula (2.5).
- **ALFM:** estimator based on an approximate (**L**ASSO) factor model as in Formula (2.8).
- **ANFM:** estimator based on an approximate (Elastic **N**et) factor model as in (2.10).

Remark 4.1 (*Latent Factor Models*)**.** Note that we also consider latent factor models in our analysis. We use the method of principal components analysis. Thereby, we impose the same structure as for observed factor models taking into account the first principal component ($K = 1$), the first three principal components ($K = 3$) and the first five principal components ($K = 5$). Additionally, we include a latent factor model based on the optimal number of factors ($K = \text{Opt}$) according to Bai and Ng (2002). We denote the latent factor models with a subscript L: EFM_L, ERFM_L, ELFM_L, ENFM_L, AFM_L, ARFM_L, ALFM_L, and ANFM_L.

Remark 4.2 (*Structure-Free Estimators*)**.** Due to completeness, we also considered the special cases with $\lambda = \infty$, thus a maximal penalty (regularization) resulting in a fully shrunk intercept-only model. The intercept-only model is a structure-free estimator of the covariance matrix as it does not consider at all the factors, or more precisely, it sets all the coefficients to zero and we end up with a covariance matrix of the residuals which in this case corresponds to the covariance matrix of the returns. Hence, for AFM we end up with the DCC-NL estimator of Engle et al. (2019) and for the EFM we end up with the diagonal of the sample covariance matrix. As it is well known that no one should use the sample covariance matrix for the estimation of large covariance matrices and that DCC-NL is outperformed by AFMs based on DCC-NL, see De Nard et al. (2021), we do not provide the corresponding results in the paper.

4.3. Main results for large dimensions

We report the following three out-of-sample performance measures for each scenario. All of them are annualized and in percent for ease of interpretation.

- **AV:** We compute the average of the 8820 out-of-sample returns and then multiply by 252 to annualize.
- **SD:** We compute the standard deviation of the 8820 out-of-sample returns and then multiply by $\sqrt{252}$ to annualize.
- **IR:** We compute the (annualized) information ratio as the ratio AV/SD.

[15] However, for $K = 1$ we estimate only the (market) factor variance by the sample variance as no shrinkage is possible.

[16] We find 43 factors, based on the revisited efficient sorting methodology and HAC standard errors that are robust against heteroskedasticity and serial correlation in the returns, that have a t-statistic larger than three.

[17] We follow here the definition of De Nard et al. (2021).

Our stance is that in the context of the GMV portfolio, the most important performance measure is the out-of-sample standard deviation, SD. The true (but unfeasible) GMV portfolio is given by (4.3). It is designed to minimize the variance (and thus the standard deviation) rather than to maximize the expected return or the information ratio. Therefore, any portfolio that implements the global minimum variance should be primarily evaluated by how successfully it achieves this goal. A high out-of-sample average return, AV, and a high out-of-sample information ratio, IR, are naturally also desirable, but should be considered of secondary importance from the point of view of evaluating the quality of a covariance matrix estimator.

We also consider the question of whether one estimation model delivers a lower out-of-sample standard deviation than another estimation model. Since we compare 9 estimation models there are 36 pairwise comparisons. To avoid a multiple testing problem and since a major goal of this paper is to show that the recommended approximate factor model is based on DCC-NL and ridge regularization, and also improves upon classical structure-free DCC-NL, we restrict attention to the comparison between the two portfolios AFM and ARFM for all factor model sizes K.[18] For a given universe and factor size, a two-sided p-value for the null hypothesis of equal standard deviations is obtained by the prewhitened HAC_{PW} method described in Ledoit and Wolf (2011, Section 3.1).[19]

The results for large portfolios and covariance matrices, $N = 1000$, are presented in Table 1 and can be summarized as follows; unless stated otherwise, the findings are with respect to the out-of-sample standard deviation as performance measure.

- All models consistently outperform $1/N$ by a wide margin.
- The approximate factor models consistently outperform the exact factor models. More specifically, the best EFM (EFM $K = 99$) has a SD of 7.50 and the worst AFM (AFM $K = 99$) has a SD of 6.90.
- Whereas for AFMs the number of factors K has no consistent effect on the estimation performance and depends on the scenario at hand, for EFMs the more factors the better. However, the 99 analyzed factors are not enough to outperform the worst case scenario of an AFM.
- For EFMs, regularization (especially ridge) can be beneficial for Fama–French factor models ($K \leq 5$).[20] For AFMs, regularization is *always* beneficial.[21]
- Moreover, ARFM consistently outperforms all other models across all number of factors. The outperformance of ARFM over AFM is always (highly) statistically significant and also economically meaningful. In general, we have the following overall ranking: ARFM, ANFM, ALFM, AFM, EFMs, $1/N$.

DeMiguel et al. (2009b) claim that it is difficult to outperform $1/N$ in terms of the out-of-sample Sharpe ratio with "sophisticated" portfolios. It can be seen that all models consistently outperform $1/N$ in terms of the out-of-sample information ratio, which translates into outperformance in terms of the out-of-sample Sharpe ratio. Note that for any factor model size K, ANFM is the best overall with an IR around two.

The outperformance of ARFM and ANFM over ALFM is not surprising. Since many of the factors have substantial correlations, asking the estimation procedure to choose between some highly-related factors for the sake of sparsity is not as good as extracting predictive information common to the factors (see, Kozak et al., 2020). To sum up, approximate factor models consistently outperform exact factor models, and ridge regularization consistently outperforms all other regularizations.

Additionally, Fig. 1 plots the out-of-sample SD surface of exact factor models and approximate (ridge) factor models for various dimensions of factors K and stocks N. It shows that ARFMs outperform exact factor models across all dimensions and that only a larger investment universe can significantly reduce the SD, whereas for exact factor models also an increase in the factor universe can improve the estimation of covariance matrices.

4.4. Latent factor models

So far we have only focused on (regularized) factor models based on the observed factor zoo. However, the models are not restricted to observed factors and we wonder how latent (regularized) factor models, based on PCA, behave in large dimensions. Therefore, we also compute the GMV portfolio for various investment universes as well as latent factor dimensions and consider the same regularization methods to compare the covariance matrix estimation performance.

The main results for the large-dimensional investment universe, $N = 1000$, are summarized in Table 2. It comes as no surprise that also for latent factors all models outperform the $1/N$ portfolio. It is worth mentioning that in general latent exact as well as latent approximate factor models benefit from regularization, especially ridge. The results are also similar in terms of the factor dimension K. Whereas for AFMs the number of factors K has only a marginal and insignificant effect on the estimation performance, for EFMs the more factors the better. However, even if we consider the optimal number of latent factors the EFMs cannot challenge the AFMs.[22]

[18] In Table 1 we see that ARFM has consistently lower SD than any other regularized approximate (and exact) factor model. Thus, we want to test if the ridge regularization significantly reduces the SD compared to no regularization.

[19] Since the out-of-sample size is very large at 8820, there is no need to use the computationally more involved bootstrap method described in Ledoit and Wolf (2011, Section 3.2), which is preferred for small sample sizes.

[20] As the small number of observed factors cannot capture all signals, regularization helps to reduce noise and extract signals.

[21] In the dynamic setting, that is when the covariance matrices are estimated with DCC, regularization is more important as it improves the out-of-sample stability of the more "volatile" estimator.

[22] The optimal number of latent factors are computed as proposed by Bai and Ng (2002). For the analyzed investment universe and period the optimal number of factors lies between four and twelve with a mean of seven. In general there is an upward trend for $K = \text{Opt}$.

Table 1
Annualized performance measures (in percent) for various estimators of the $N = 1000$ GMV portfolio. AV stands for average return; SD stands for standard deviation; and IR stands for information ratio. All measures are based on 8820 daily out-of-sample returns from 01/01/1982 until 12/31/2016. In the rows labeled SD, the lowest number appears in **bold face**. In the column labeled ARFM, significant outperformance over the AFM in terms of SD is denoted by asterisks: *** denotes significance at the 0.01 level; ** denotes significance at the 0.05 level; and * denotes significance at the 0.1 level.

N = 1000									
	1/N	EFM	ERFM	ELFM	ENFM	AFM	ARFM	ALFM	ANFM
K = 1									
AV	13.975	13.493	13.167	13.526	13.160	13.681	12.363	13.641	13.635
SD	17.427	11.303	9.806	10.974	10.709	6.847	**6.482*****	6.779	6.808
IR	0.802	1.194	1.343	1.233	1.229	1.998	1.907	2.012	2.003
K = 3									
AV	13.975	13.375	13.369	13.808	13.162	13.257	12.543	13.350	13.370
SD	17.427	10.579	9.351	10.165	10.164	6.814	**6.524*****	6.836	6.802
IR	0.802	1.264	1.430	1.358	1.295	1.946	1.923	1.953	1.965
K = 5									
AV	13.975	13.735	13.317	14.154	13.217	13.091	12.707	13.230	13.474
SD	17.427	10.044	9.186	9.686	9.894	6.834	**6.524*****	6.836	6.784
IR	0.802	1.368	1.450	1.461	1.336	1.916	1.948	1.935	1.986
K = 48									
AV	13.975	13.516	13.773	14.342	13.501	12.900	12.994	13.163	13.657
SD	17.427	7.777	8.811	8.295	9.278	6.876	**6.533*****	6.840	6.766
IR	0.802	1.738	1.563	1.729	1.455	1.876	1.989	1.925	2.019
K = 99									
AV	13.975	13.332	13.494	14.343	13.738	12.898	13.133	13.142	13.652
SD	17.427	7.503	8.700	8.186	9.312	6.900	**6.593*****	6.868	6.758
IR	0.802	1.777	1.551	1.752	1.475	1.869	1.992	1.913	2.020

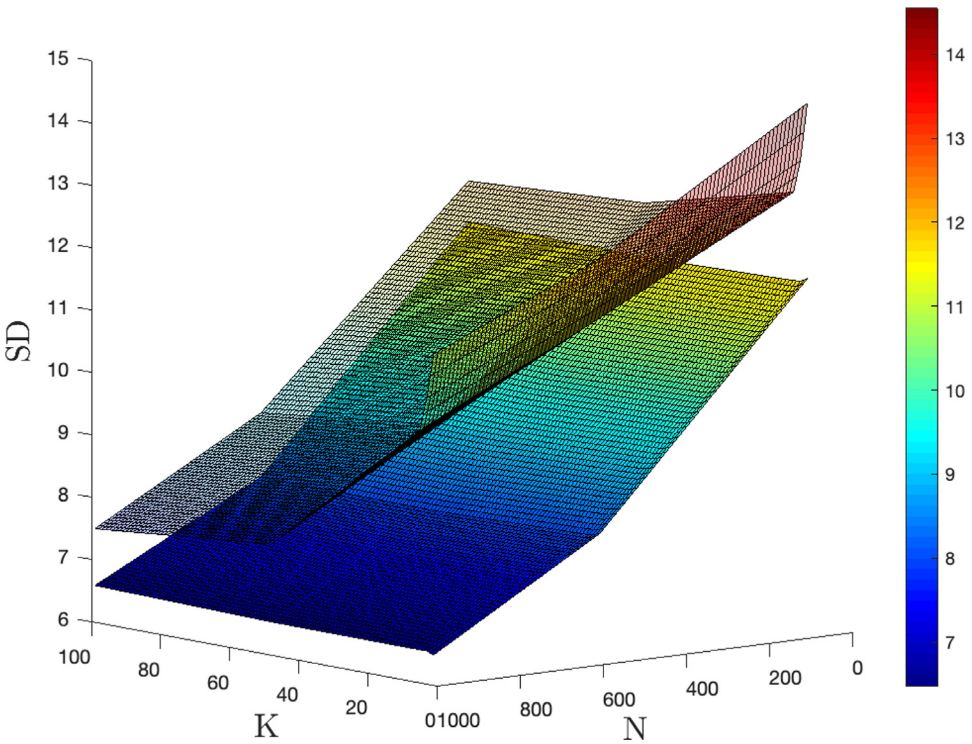

Fig. 1. SD of the EFM (upper transparent surface) and ARFM (non transparent lower surface) for various number of assets and factors.

Table 2
Annualized performance measures (in percent) for various (latent) estimators of the $N = 1000$ GMV portfolio. AV stands for average return; SD stands for standard deviation; and IR stands for information ratio. All measures are based on 8820 daily out-of-sample returns from 01/01/1982 until 12/31/2016. In the rows labeled SD, the lowest number appears in **bold face**. In the column labeled $\mathrm{ARFM_L}$, significant outperformance over the $\mathrm{AFM_L}$ in terms of SD is denoted by asterisks: *** denotes significance at the 0.01 level; ** denotes significance at the 0.05 level; and * denotes significance at the 0.1 level.

N = 1000								
	$\mathrm{EFM_L}$	$\mathrm{ERFM_L}$	$\mathrm{ELFM_L}$	$\mathrm{ENFM_L}$	$\mathrm{AFM_L}$	$\mathrm{ARFM_L}$	$\mathrm{ALFM_L}$	$\mathrm{ANFM_L}$
K = 1								
AV	13.054	12.890	13.082	13.040	13.704	12.287	13.651	13.727
SD	11.352	9.564	11.084	10.316	6.886	**6.501***	6.822	6.856
IR	1.150	1.348	1.180	1.264	1.990	1.890	2.001	2.002
K = 3								
AV	14.160	13.870	14.178	13.153	13.297	12.243	13.445	13.775
SD	10.075	9.053	9.558	9.666	6.876	**6.435***	6.810	6.726
IR	1.405	1.532	1.483	1.361	1.934	1.903	1.974	2.048
K = 5								
AV	13.082	13.205	13.334	13.147	12.684	11.946	12.820	13.777
SD	8.401	8.625	8.104	9.567	6.875	**6.405***	6.835	6.714
IR	1.557	1.531	1.645	1.374	1.845	1.865	1.876	2.052
K = Opt								
AV	13.320	13.543	13.557	13.379	12.825	12.155	12.921	12.879
SD	7.606	8.482	7.440	7.995	6.843	**6.420***	6.848	6.947
IR	1.751	1.597	1.822	1.673	1.874	1.893	1.887	1.854

Consequently, the approximate factor models consistently and markedly outperform the exact factor models. Again, across all scenarios the ARFM is best overall and significantly reduces the SD. Furthermore, if we compare the results with the observed factors (Table 1) we see that the out-of-sample standard deviation is on a similar level, however, the latent factor models perform (slightly) better.

The outperformance of latent factor models basically means that there is no need to laboriously compute the (observed) factor zoo, as it is enough to focus on the information in the underlying return time series and simply compute the latent factors. The first K = Opt latent factors should then be used in an approximate factor model with ridge regularization. This is also supported by various robustness checks, presented in the next section, and the supplementary material. Note that the Bai and Ng (2002) criterion we use tends to overestimate the number of factors. However, since we still use ridge in latent factor models, the potential overestimation is not an issue as the spurious ones are probably shrunk to zero. In sum, the best performing model, across all portfolio sizes (N) and number of factors (K), is the latent ARFM model with the optimal K selection.

4.5. Robustness checks

In the supplementary material we detail and present various robustness checks.

We find that for EFMs, regularization (especially ridge) can be beneficial for smaller dimensions of N and/or K. For AFMs, regularization is *always* beneficial. Moreover, ARFM consistently outperforms all other models across all portfolio sizes and numbers of factors. The outperformance of ARFM over AFM is always statistically significant and also economically meaningful. Consequently, the results on the factor model structure, size and regularization method are robust in terms of investment universe dimension.

It might be natural to ask whether the relative performance of the various models is stable during that period or whether it changes during certain subperiods, such as periods of "boom" vs. periods of "bust". To answer this question, we carry out a rolling-window analysis based on shorter out-of-sample periods: one month (21 days), one year (252 days), and five years (1260 days). We find that the relative performance is remarkably stable over time and that, in particular, ARFM generally performs the best during all subperiods and shorter evaluation lengths.

The results presented so far are for the new estimation scheme introduced in Section 3. When we compare the results, we see that the new estimation scheme consistently outperforms the traditional estimation scheme across all factor models, as well as across all investment universes and factor dimensions. Even though the outperformance is usually marginal and not (always) significant, we recommend updating the estimation scheme, as averaging portfolio weights clearly dominates the previous approach of averaging forecasted covariance matrices. Additionally, the findings on the traditional estimation scheme give further support that approximate factor models perform better than exact factor models across all scenarios and that the ridge regularization, together with DCC-NL, seems to be the most useful for covariance matrix estimation.

We also analyze how stable and diversified the portfolios are, looking at the average turnover and leverage. Across all dimensions and regularization methods, exact factor models have lower turnover and leverage than approximate factor models. The portfolios of the new averaging scheme look similar to the traditional approach, but it is worth mentioning that averaging portfolios instead of covariance matrices, always reduces the turnover and leverage.

Remember that our primary goal is to improve upon the estimation of large-dimensional covariance matrices, namely, second moments, and not upon the estimation of first moments. Nevertheless, to further prove the robustness of our results and to see how our estimators perform in a more realistic scenario for benchmarked managers, we analyze a "full" Markowitz portfolio with a signal. For a Markowitz portfolio with momentum signal, the approximate factor models also consistently outperform the exact factor models. More specifically, the best EFM (EFM with $K = 99$) has an IR of 1.73 and the worst AFM (AFM $K = 99$) has an IR of 1.81. Whereas for AFMs the number of factors K has no consistent effect on the IR and depends on the scenario at hand, for EFMs the more factors the better. However, the 99 analyzed factors are not enough to outperform the worst case scenario of an AFM. ANFM consistently outperforms all other models across all number of factors. Engle and Colacito (2006) argue for the use of the out-of-sample standard deviation, SD, as a performance measure also in the context of a "full" Markowitz portfolio. For this alternative performance, ARFM also has the lowest out-of-sample standard deviation. The outperformance is markedly and always statistically significant. Finally, the results are consistent with the net of transaction costs IR. Even though EFMs have lower turnover than AFMs, for a transaction costs of 10 bps, AFMs still (significantly) outperform EFMs.

Finally, we also analyze static exact and approximate factor models as defined in De Nard et al. (2021). Hence, the factor and residual covariance matrices are estimated without any GARCH model, but with nonlinear shrinkage which is a static estimator. We support the results of De Nard et al. (2021) that dynamic models consistently and often markedly outperform static models. In general we observe the following ranking: (i) dynamic approximate factor models, (ii) static approximate factor models, and (iii) exact factor models. The results show that both, nonlinear shrinkage and DCC models help to improve portfolio performance and thus derive better covariance matrix estimators.

4.6. Summary of results

We have carried out an extensive backtest analysis, evaluating the out-of-sample performance of our dynamic double-shrinkage (approximate) factor model based on a new DCC-NL estimation scheme. Specifically, we have compared ARFM to a number of other strategies – various factor models and structure-free estimators of the covariance matrix – to estimate the global minimum-variance portfolio and the Markowitz portfolio with momentum signal.

First, we find that the new DCC-NL estimation scheme consistently outperforms the previous estimation scheme across all factor models, as well as across all investment universe sizes and factor dimensions. Even though the outperformance is usually marginal and not (always) significant, we recommend updating the estimation scheme, as averaging portfolio weights clearly dominates the previous approach of averaging forecasted covariance matrices.

Second, among the considered portfolios, ARFM is the clear winner. In most scenarios, ARFM performs the best, followed by ANFM, ALFM, AFM, EFMs, 1/N. We show that for AFMs, regularization is always beneficial and thus we recommend updating the previous AFM-DCC-NL of De Nard et al. (2021) with the double-shrinkage estimation. Additionally, the outperformance of ARFM over AFM is always statistically significant and also economically meaningful, giving even statistical evidence for the benefit of shrinking factors. Whereas for AFMs the number of factors K has no consistent effect on the estimation performance and depends on the scenario at hand (N, K and AFM version), for EFMs the more factors the better. However, the 99 observed factors are not enough to outperform the worst case scenario of an AFM. Therefore, AFMs consistently outperform EFMs, suggesting the importance of the second shrinkage. Note that we run several robustness checks to confirm the results across various portfolios sizes, numbers of factors, subperiods and shorter evaluation lengths.

Consequently, we confirm the results of De Nard et al. (2021) that including multiple factors (using (regularized) approximate multi-factor models) does not necessarily result in better performance; on the contrary, doing so can actually reduce the performance due to the additional estimation uncertainty. The main lesson is that the market factor is too outsized to be ignored, even by estimators that draw from state-of-the-art techniques in large-dimensional asymptotics and conditional heteroskedasticity; on the other hand, additional factors do not seem to be needed for sophisticated estimation methods.

In terms of portfolio turnover and leverage, we find that EFMs perform markedly better than AFMs. However, AFMs still have higher net-of-transaction-costs information ratios for Markowitz portfolios with momentum signal.

Finally, we show that dynamic double-shrinkage factor models perform even better with latent factors based on PCA. Hence, we recommend avoiding the factor zoo and focusing on the information in the underlying return time series and simply computing the optimal number of latent factors as proposed by Bai and Ng (2002). In sum, the ARFM_L performs the best.

5. Conclusion

This paper reconciles a traditional feature of covariance matrix estimation in finance, namely, factor models, with more modern methods based on large-dimensional asymptotic theory. We demonstrate on historical data that there is a net benefit in combining these two approaches. The key is to allow for conditional heteroskedasticity, together with nonlinear shrinkage estimation of the factor model coefficients (regularization) as well as nonlinear shrinkage estimation of the resulting large-dimensional residual and factor covariance matrices, as in our new $\text{ARFM}_{(\text{L})}$ model. As a secondary contribution, we propose an improved scheme for extrapolating the covariance matrix forecast over the holding period of the investment strategy in case the holding period (monthly) exceeds the frequency of the observed data (daily). More specifically, instead of averaging forecasted covariance matrices, we suggest averaging forecasted portfolio weights.

Finally, we find that a structured estimator with shrinkage techniques, such as ridge and DCC-NL, significantly improves estimation accuracy of covariance matrices and should be considered across all dimensions. Nonetheless, as exact factor models are consistently and markedly outperformed by approximate factor models, and as approximate factor models do not profit from a large number of factors, there is no need for laboriously taking into account the entire factor zoo. On the contrary, we argue that it is enough to use the Fama–French factors, or even only the market factor, in an approximate factor model based on ridge regularization and DCC-NL. This finding makes the resulting ARFM estimator even more attractive for portfolio managers, since only data on the market factor are needed. Additionally, we find that the proposed double-shrinkage estimator works even better for latent factor models. Consequently, we suggest focusing only on the information in the underlying return time series and simply computing the (optimal number of) latent factors for the $\mathrm{ARFM_L}$ model. In short, for the estimation of covariance matrices, taming the factor zoo is clearly beneficial, however, avoiding it is even better.

Taken together, these techniques should help portfolio managers develop better-performing investment strategies, and should also help empirical finance academics develop more powerful predictive tests for anomalies in the cross-section of stock returns.

CRediT authorship contribution statement

Gianluca De Nard: Conceptualization, Methodology, Formal analysis, Resources, Writing – original draft, Supervision, Project administration. **Zhao Zhao:** Software, Validation, Investigation, Data curation, Writing – review & editing, Visualization, Funding acquisition.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jempfin.2023.02.003.

Factor Computation: A detailed description on how to compute the 94 analyzed (common) factors based on efficient sorting.

Estimation Algorithms for Regularization: Details on the accelerated proximal gradient algorithm for the estimation of regularization parameters.

Forecasting Covariance Matrices: Details on how the ‘‘variance’’ prediction is performed.

Sample Splitting and Tuning via Validation: Details on the training and validation framework.

Robustness Checks: Additional results and various robustness checks on investment-universe dimension, sub-period analysis, portfolio-averaging scheme, turnover and leverage, Markowitz portfolio with momentum signal, and static factor models.

References

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70 (1), 191–221.

Bai, J., Shi, S., 2011. Estimating high dimensional covariance matrices and its applications. Ann. Econ. Finance 12 (2), 199–215.

Bollerslev, T., Patton, A.J., Quaedvlieg, R., 2018. Modeling and forecasting (un) reliable realized covariances for more reliable financial decisions. J. Econometrics 207 (1), 71–91.

Chincarini, L.B., Kim, D., 2006. Quantitative Equity Portfolio Management: An Active Approach to Portfolio Construction and Management. McGraw-Hill, New York.

Cochrane, J.H., 2011. Presidential address: Discount rates. J. Finance 66 (4), 1047–1108.

Cremers, M., Petajisto, A., Zitzewitz, E., 2012. Should Benchmark Indices Have Alpha? Revisiting Performance Evaluation. Nber working paper 18050, National Bureau of Economic Research.

De Nard, G., 2022. Oops! I shrunk the sample covariance matrix again: Blockbuster meets shrinkage. J. Financ. Econom. 20 (4), 569–611.

De Nard, G., Engle, R.F., Ledoit, O., Wolf, M., 2022. Large dynamic covariance matrices: Enhancements based on intraday data. J. Bank. Financ. 138.

De Nard, G., Ledoit, O., Wolf, M., 2021. Factor models for portfolio selection in large dimensions: The good, the better and the Ugly. J. Financ. Econom. 19 (2), 236–257.

DeMiguel, V., Garlappi, L., Nogales, F.J., Uppal, R., 2009a. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. Manage. Sci. 55 (5), 798–812.

DeMiguel, V., Garlappi, L., Uppal, R., 2009b. Optimal versus naive diversification: How inefficient is the $1/N$ portfolio strategy? Rev. Financ. Stud. 22, 1915–1953.

DeMiguel, V., Martin-Utrera, A., Nogales, F.J., 2013. Size matters: Optimal calibration of shrinkage estimators for for portfolio selection. J. Bank. Financ. 37, 3018–3034.

Engle, R.F., Colacito, R., 2006. Testing and valuing dynamic correlations for asset allocation. J. Bus. Econom. Statist. 24 (2), 238–253.

Engle, R.F., Ledoit, O., Wolf, M., 2019. Large dynamic covariance matrices. J. Bus. Econom. Statist. 37, 363–375.

Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. J. Finance 47 (2), 427–465.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. J. Financ. Econ. 33 (1), 3–56.

Fama, E., French, K., 2008. Dissecting anomalies. J. Finance LXIII(4), 1653–1678.

Fama, E.F., French, K.R., 2015. A five factor asset pricing model. J. Financ. Econ. 116 (1), 1–22.

Fan, J., Fan, Y., Lv, J., 2008. High dimensional covariance matrix estimation using a factor model. J. Econometrics 147 (1), 186–197.

Fan, J., Kim, D., 2018. Robust high-dimensional volatility matrix estimation for high-frequency factor model. J. Amer. Statist. Assoc. 113 (523), 1268–1283.

Fan, J., Kim, D., 2019. Structured volatility matrix estimation for non-synchronized high-frequency financial data. J. Econometrics 209 (1), 61–78.

Fan, J., Liao, Y., Liu, H., 2016. An overview of the estimation of large covariance and precision matrices. Econom. J. 19, C1–C32.

Fan, J., Liao, Y., Mincheva, M., 2013. Large covariance estimation by thresholding principal orthogonal complements (with discussion). J. R. Stat. Soc. Ser. B Stat. Methodol. 75 (4), 603–680.

Feng, G., Giglio, S., Xiu, D., 2020. Taming the factor zoo: A test of new factors. J. Finance 75 (3), 1327–1370.

Freyberger, J., Neuhierl, A., Weber, M., 2020. Dissecting characteristics nonparametrically. Rev. Financ. Stud. 33 (5), 2326–2377.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. Rev. Financ. Stud. 33 (5), 2223–2273.
Gu, S., Kelly, B., Xiu, D., 2021. Autoencoder asset pricing models. J. Econometrics 222 (1), 429–450.
Han, C., 2020. How much should portfolios shrink? Financial Manag. 49 (3), 707–740.
Han, Y., Huang, D., Zhou, G., 2021. Anomalies enhanced: A portfolio rebalancing approach. Financial Manag. 50 (2), 371–424.
Harvey, C.R., Liu, Y., 2015. Backtesting. J. Portfolio Manag. 42 (1), 13–28.
Haugen, R.A., Baker, N.L., 1991. The efficient market inefficiency of capitalization–weighted stock portfolios. J. Portf. Manag. 17 (3), 35–40.
Hou, K., Xue, C., Zhang, L., 2020. Replicating anomalies. Rev. Financ. Stud. 33 (5), 2019–2133.
Jagannathan, R., Ma, T., 2003. Risk reduction in large portfolios: Why imposing the wrong constraints helps. J. Finance 54 (4), 1651–1684.
Kelly, B.T., Pruitt, S., Su, Y., 2019. Characteristics are covariances: A unified model of risk and return. J. Financ. Econ. 134 (3), 501–524.
Kim, D., Fan, J., 2019. Factor GARCH-Itô models for high-frequency data with application to large volatility matrix prediction. J. Econometrics 208 (2), 395–417.
Kozak, S., Nagel, S., Santosh, S., 2020. Shrinking the cross-section. J. Financ. Econ. 135 (2), 271–292.
Ledoit, O., Wolf, M., 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. J. Empir. Financ. 10 (5), 603–621.
Ledoit, O., Wolf, M., 2011. Robust performance hypothesis testing with the variance. Wilmott Mag. September, 86–89.
Ledoit, O., Wolf, M., 2015. Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. J. Multivariate Anal. 139 (2), 360–384.
Ledoit, O., Wolf, M., 2017. Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. Rev. Financ. Stud. 30 (12), 4349–4388.
Ledoit, O., Wolf, M., 2020. Analytical nonlinear shrinkage of large-dimensional covariance matrices. Ann. Statist. 48 (5), 3043–3065.
Ledoit, O., Wolf, M., Zhao, Z., 2019. Efficient sorting: A more powerful test for cross-sectional anomalies. J. Financ. Econom. 17 (4), 645–686.
Lioui, A., Tarelli, A., 2020. Factor investing for the long run. J. Econom. Dynam. Control 117, 103960.
McLean, R.D., Pontiff, J., 2016. Does academic research destroy stock return predictability? J. Finance 71 (1), 5–32.
Meucci, A., 2005. Risk and Asset Allocation. Springer Finance, Berlin Heidelberg New York.
Nielsen, F., Aylursubramanian, R., 2008. Far From the Madding Crowd — Volatility Efficient Indices. Research insights, MSCI Barra.
Pakel, C., Shephard, N., Sheppard, K., Engle, R.F., 2021. Fitting vast dimensional time-varying covariance models. J. Bus. Econom. Statist. 39 (3), 652–668.
Ross, S.A., 1976. The arbitrage theory of capital asset pricing. J. Econom. Theory 13 (3), 341–360.
Sharpe, W.F., 1963. A simplified model for portfolio analysis. Manage. Sci. 9 (1), 277–293.
So, M.K., Chan, T.W., Chu, A.M., 2020. Efficient estimation of high-dimensional dynamic covariance by risk factor mapping: Applications for financial risk management. J. Econometrics.
Sun, Y., Xu, W., 2021. A factor-based estimation of integrated covariance matrix with noisy high-frequency data. J. Bus. Econom. Statist. 1–15.