University of Zurich

Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 290

Factor Models for Portfolio Selection in Large Dimensions: The Good, the Better and the Ugly

Gianluca De Nard, Olivier Ledoit and Michael Wolf

First version: June 2018
This version: December 2018

Factor Models for Portfolio Selection in Large Dimensions: The Good, the Better and the Ugly

Gianluca De Nard
Department of Banking and Finance
University of Zurich
CH-8032 Zurich, Switzerland
gianluca.denard@bf.uzh.ch

Olivier Ledoit
Department of Economics
University of Zurich
CH-8032 Zurich, Switzerland
olivier.ledoit@econ.uzh.ch

Michael Wolf*
Department of Economics
University of Zurich
CH-8032 Zurich, Switzerland
michael.wolf@econ.uzh.ch

First version: June 2018
This version: December 2018

Abstract

This paper injects factor structure into the estimation of time-varying, large-dimensional covariance matrices of stock returns. Existing factor models struggle to model the covariance matrix of residuals in the presence of time-varying conditional heteroskedasticity in large universes. Conversely, rotation-equivariant estimators of large-dimensional time-varying covariance matrices forsake directional information embedded in market-wide risk factors. We introduce a new covariance matrix estimator that blends factor structure with time-varying conditional heteroskedasticity of residuals in large dimensions up to 1000 stocks. It displays superior all-around performance on historical data against a variety of state-of-the-art competitors, including static factor models, exogenous factor models, sparsity-based models, and structure-free dynamic models. This new estimator can be used to deliver more efficient portfolio selection and detection of anomalies in the cross-section of stock returns.

KEY WORDS: Dynamic conditional correlations, factor models, multivariate GARCH, Markowitz portfolio selection, nonlinear shrinkage.

JEL CLASSIFICATION NOS: C13, C58, G11.

*Corresponding author. Address: Zürichbergstrasse 14, 8032 Zurich, Switzerland. Phone: +41 44 634 5096.

1 Introduction

Factor models have a long history in finance, with wide range of applications both in theory and in practice. Examples of theoretical applications are asset pricing models, such as the Capital Asset Pricing Model (CAPM) and the Arbitrage Pricing Theory (APT) of Ross (1976), and various fund-separation theorems. In practice, factor models are used, among others, to evaluate the performance of portfolio managers, to assess return anomalies, to predict returns, and to construct portfolios; for example, see Meucci (2005) and Chincarini and Kim (2006).

The aim of this paper is to study the usefulness of factor models for the estimation of covariance matrices in large dimensions, with the goal of using such matrices as input for Markowitz portfolio selection. The literature on Markowitz portfolio selection in large dimensions — that is, when the number of assets is of the same magnitude as the number of observations — has experienced a large expansion over the last decade or so and is still growing; 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), and Ledoit and Wolf (2017).

Markowitz portfolio selection, in its general form, requires estimates of (i) the vector of expected returns and (ii) the covariance matrix of the assets in the investment universe. Green et al. (2013) list over 300 papers that address the first estimation problem, and it can be assumed that many more such papers have been written in the meantime. On the other hand, we only aim to address the second problem, the estimation of the covariance matrix.

In particular, we want to study the usefulness of factor models to this end and address the following questions, among others. Is it better to use a static factor model (which assumes a time-invariant covariance matrix) or a dynamic factor model (which assumes a time-varying covariance matrix)? Is it better to use observed factors (such as Fama-French factors) or latent factors estimated from past return data (such as principal components)? How much, if any, do approximate factor models (AFMs) improve over an exact factor model (EFM)? Is it better to use several factors or maybe just a single market factor? And, last but not least, do factor models improve at all over (sophisticated) structure-free estimators of the covariance matrix? (Note that we shall ask these questions only for unconditional factor models, where the factor loadings are time-invariant. This is because we have found, in unreported results, that conditional factor models, where the factor loadings are time-varying, tend to give worse results in the context of portfolio selection.)

Since the answers to these questions can be assumed to depend on the context — that is, the number and the nature of the assets as well as the frequency of the observed returns — we only consider one specific context where the assets are individual stocks and the return frequency is daily. Arguably, this is the context of most interest to real-life portfolio managers.

It should furthermore be clear that, in the absence of an overly strict and thus unrealistic model for asset returns, these questions cannot be answered theoretically, or even by a Monte Carlo study, and must therefore be addressed using back-tests on historical data. We use a large data set on US stock returns dating back 45 years.

A further contribution of this paper is to propose a new forecasting scheme for covariance matrices in the case where a dynamic 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, we propose an 'averaged forecast' of the covariance matrix by averaging over l-steps-ahead forecasts (at the frequency of the observed returns) over all periods l in the holding period.

The remainder of the paper is organized as follows. Section 2 gives a brief description of static and dynamic factor models. Section 3 details our new 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 gives an outlook on future research. Section 6 concludes. Appendix A contains all the tables.

2 Factor Models

2.1 Notation

In what follows, the subscript $i \in \{1, \dots, N\}$ indexes the assets, where N denotes the dimension of the investment universe; the subscript $k \in \{1, \dots, K\}$ indexes the (common) factors, where K denotes the number of factors; and the subscript $t \in \{1, \dots, T\}$ indexes the dates, where T denotes the sample size. The notation $\mathsf{Corr}(\cdot)$ 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 $r_t := (r_{1,t}, \dots, r_{N,t})'$
- $f_{k,t}$: return for factor k at date t, stacked into $f_t := (f_{1,t}, \dots, f_{K,t})'$
- $u_{i,t}$: error term for asset i at date t, stacked into $u_t := (u_{1,t}, \dots, u_{N,t})'$
- $x_{i,t}$: underlying time series for covariance matrix estimation; thus, $x_{i,t} \in \{r_{i,t}, u_{i,t}\}$
- $d_{i,t}^2 := \mathsf{Var}(x_{i,t}|\mathcal{F}_{t-1})$: conditional variance of the ith variable at date t
- $s_{i,t} := x_{i,t}/d_{i,t}$: devolatized series, stacked into $s_t := (s_{1,t}, ..., s_{N,t})'$
- D_t: the N-dimensional diagonal matrix whose ith diagonal element is $d_{i,t}$
- $R_t := \mathsf{Corr}(x_t|\mathcal{F}_{t-1}) = \mathsf{Cov}(s_t|\mathcal{F}_{t-1})$: conditional correlation matrix at date t
- $\Sigma_t := \mathsf{Cov}(x_t|\mathcal{F}_{t-1})$: conditional covariance matrix at date t; thus $\mathsf{Diag}(\Sigma_t) = D_t^2$
- $C := \mathbb{E}(R_t) = \mathsf{Corr}(x_t) = \mathsf{Cov}(s_t)$: unconditional correlation matrix
- $\Sigma_{f,t} := \mathsf{Cov}(f_t|\mathcal{F}_{t-1})$: conditional covariance matrix of the factors

Here, the symbol $:=$ is a definition sign, where the left-hand side is defined to be equal to the right-hand side whereas the symbol $\equiv$ denotes that the left-hand side is constantly equal to the right-hand side.

In our description below, the factors are allowed to be observed, such as Fama-French factors, or latent (and thus estimated from historical return data in a prior step), such as principal components.

2.2 Static Factor Models

A static factor model assumes that, for every asset $i = 1, \ldots, N$,

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

with $\beta_i := (\beta_{i,1}, \ldots, \beta_{i,K})'$ and $\mathbb{E}(u_{i,t}|f_t) = 0$. Furthermore, it is assumed that $\Sigma_{f,t} \equiv \Sigma_f$ and $\Sigma_{u,t} \equiv \Sigma_u$, for all $t = 1, \ldots, T$.

An alternative formulation is based on risk-free returns $r_{i,t} - r_{f,t}$, where $r_{f,t}$ denotes the return of the risk-free asset in period t, instead of raw returns $r_{i,t}$; in such a case it is common to exclude the intercept α_i in model (2.1).

The key assumptions that make a factor model a static one are: (a) the intercepts α_i and the factor loadings β_i are time-invariant; (b) the conditional covariance matrix of the vector of factors f_t is time-invariant; and (c) the conditional covariance matrix of the vector of errors u_t is time-invariant. There may be other definitions for a static factor model, but this is the one that we shall adopt for the purpose of this paper.

An *exact* factor model assumes in addition that Σ_u is a diagonal matrix. In contrast, an *approximate* factor model only assumes that Σ_u is a matrix with bounded L^1 or L^2 norm; for example, see Connor and Korajczyk (1993), Bai and Ng (2002), Fan et al. (2008), and the references therein.

There is an important distinction between observed and latent factors. Observed factors are known and based on outside information. The leading example, and the one to which we will restrict attention, is the one of Fama-French factors. But many other factors have been proposed in the literature; for example, see Bai and Shi (2011, Section 4) and Feng et al. (2017). 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 and principal components; for example, see Bai and Shi (2011, Section 5). In our analysis, we shall use the method of principal components which, under suitable regularity conditions, leads to consistent estimation of the factors (up to a rotation) and we tacitly assume such a set of conditions; for example, see Fan et al. (2013, Section 2).

Collect the time-invariant factor loadings across assets into a $K \times N$ matrix B, where β_i is the ith column of B. Under the model assumptions, the time-invariant covariance matrix of r_t

is then given by

$$\Sigma_r = B'\Sigma_f B + \Sigma_u \ . \tag{2.2}$$

The estimation of a static factor model is straightforward. The intercepts and the factor loadings can be estimated by ordinary least squares (OLS), for $i = 1, \ldots, N$, resulting in residuals $\hat{u}_t$, for $t = 1, \ldots, T$. The covariance matrix Σ_f can be estimated by the sample covariance matrix of the $\{f_t\}$; denote the resulting estimator by $\hat{\Sigma}_f$. The estimation of the covariance matrix Σ_u depends on whether an exact or an approximate factor model is assumed; in either case, the starting point is the sample covariance matrix of the residuals $\{\hat{u}_t\}$, denoted by $S_{\hat{u}}$. In the case of an exact factor model, it is customary to take $\mathsf{Diag}(S_{\hat{u}})$ as the estimator of Σ_u. In the case of an approximate factor model, one can apply thresholding to $S_{\hat{u}}$ in order to arrive at a sparse estimator; for example, see Fan et al. (2013) for some proposals to this end. In either case, denote the estimator of Σ_u by $\hat{\Sigma}_u$.

The estimator of Σ_r is then given by

$$\hat{\Sigma}_r := \hat{B}'\hat{\Sigma}_f \hat{B} + \hat{\Sigma}_u \ , \tag{2.3}$$

where $\hat{B}$ is a $K \times N$ matrix whose ith column is the vector $\hat{\beta}_i$.

2.3 Dynamic Factor Models

Recall the three assumptions concerning model (2.1) that make a factor model static, at least according to our definition: (a) the intercepts and the factor loadings are time-invariant; (b) the conditional covariance matrix of the vector of factors f_t is time-invariant; and (c) the conditional covariance matrix of the vector of errors u_t is time-invariant.

A dynamic factor model is then one in which at least one of the three assumptions is violated (though not necessarily all three together). As a consequence, at least one of the following three generalizations must be in place: (a') the intercepts α_i and the factor loadings β_i are allowed to be time-varying; (b') the conditional covariance matrix of the vector of factors f_t is allowed to be time-varying; or (c') the conditional covariance matrix of the vector of errors u_t is allowed to be time-varying. There are certainly other definitions of a dynamic factor model — for example, see Stock and Watson (2011) and the references therein — but this is the one that we shall adopt for the purpose of this paper.

Generalization (a'), allowing the intercepts and the factor loadings to be time-varying, results in what is commonly referred to as a *conditional* factor model; for example, see Avramov and Chordia (2006), Ang and Kristensen (2012), Engle (2016) and Bali et al. (2017). Such models often work better than *unconditional* factor models (where the intercepts and the factor loadings are time-invariant) in the context of asset pricing. But in empirical results not reported here, we have found this not to be true in the context of portfolio selection; therefore, we will not consider conditional factor models in the remainder of this paper.

Moving on to generalizations (b') and (c'), an unconditional dynamic factor model also assumes that, for every asset $i = 1, \ldots, N$,

$$r_{i,t} = \alpha_i + \beta_i' f_t + u_{i,t} \ , \tag{2.4}$$

with $\beta_i := (\beta_{i,1}, \ldots, \beta_{i,K})'$ and $\mathbb{E}(u_{i,t}|f_t) = 0$. But now the conditional covariance matrices of f_t and u_t may both be time-varying; recall the corresponding notations defined in Section 2.1.

Under the model assumptions, the time-varying conditional covariance matrix of r_t is then given by

$$\Sigma_{r,t} = B'\Sigma_{f,t}B + \Sigma_{u,t} \ , \tag{2.5}$$

where B is again the $K \times N$ matrix whose ith column is the vector β_i.

The estimation of an unconditional dynamic factor model starts in the same way as the estimation of a static factor model: estimate the time-invariant intercepts and the factor loadings by OLS, which also yields the residuals $\hat{u}_t$. Furthermore, denote the estimators of $\Sigma_{f,t}$ and $\Sigma_{u,t}$ by $\hat{\Sigma}_{f,t}$ and $\hat{\Sigma}_{u,t}$, respectively; corresponding estimators will be discussed in Section 3.2.

The estimator of $\Sigma_{r,t}$ is then given by

$$\hat{\Sigma}_{r,t} := \hat{B}\hat{\Sigma}_{f,t}\hat{B} + \hat{\Sigma}_{u,t} \ . \tag{2.6}$$

Note that, as a special case, either $\Sigma_{f,t}$ or $\Sigma_{u,t}$, but not both, may be assumed to be time-invariant and then estimated in a fashion that would be appropriate for the static models of Section 2.2.

If $\Sigma_{f,t}$ is assumed to be time-invariant, the estimator of $\Sigma_{r,t}$ is then given by

$$\hat{\Sigma}_{r,t} := \hat{B}\hat{\Sigma}_{f}\hat{B} + \hat{\Sigma}_{u,t} \ . \tag{2.7}$$

On the other hand, if $\Sigma_{u,t}$ is assumed to be time-invariant, the estimator of $\Sigma_{r,t}$ is given by

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

3 New Estimation Schemes

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.

3.1 Static Factor Models

In line with most of the literature these days, we believe that the assumption of an exact factor model is overly strict in practice and hence favor approximate factor models instead (although we will also include an exact factor model for reference in the empirical analysis of Section 4).

Based on the nature of approximate factor models, which assume that the covariance matrix Σ_u is sparse or with bounded eigenvalues, one natural candidate is to use an estimator of Σ_u that imposes sparsity; for example, see Fan et al. (2013) for some proposals to this end. All these proposals apply some thresholding scheme to $S_{\hat{u}}$, the sample covariance matrix of the residuals $\{\hat{u}_t\}$.

As an alternative, we propose to apply the nonlinear shrinkage method of Ledoit and Wolf (2017) to the matrix $S_{\hat{u}}$. This nonlinear shrinkage method does not impose any structure on the estimator and, in particular, will not necessarily result in a sparse estimator. However, this feature may actually be an advantage rather than a disadvantage, with the reasoning being as follows.

When the assumption of an approximate factor model is true, nonlinear shrinkage of $S_{\hat{u}}$ will at least result in a sparser estimator compared to $S_{\hat{u}}$, since nonlinear shrinkage will deliver an estimator that is closer to the truth compared to the sample covariance matrix; this has been demonstrated by large-dimensional asymptotic theory, extensive Monte Carlo experiments, and empirical applications in Ledoit and Wolf (2012, 2015, 2017). As a consequence, when the assumption of an approximate factor model is true, applying nonlinear shrinkage to $S_{\hat{u}}$ might perform about as well as applying a thresholding scheme to $S_{\hat{u}}$.

On the other hand, when the assumption of an approximate factor model is false — that is, when the factors used are not really factors or only 'weak' factors — then the assumption of sparsity of Σ_u does not hold true. In such a case, applying nonlinear shrinkage to $S_{\hat{u}}$ should perform better than applying a thresholding scheme to $S_{\hat{u}}$.

By this reasoning, applying nonlinear shrinkage to $S_{\hat{u}}$ may be more robust to the number (and nature) of factors used than applying a thresholding scheme that enforces sparsity.

3.2 Dynamic Factor Models

Based on formulas (2.6)–(2.8), what is needed in addition to the estimation of a static factor model are estimators $\hat{\Sigma}_{f,t}$ or $\hat{\Sigma}_{u,t}$. We propose to use a multivariate GARCH model to this end; in particular, we recommend the dynamic conditional correlation (DCC) model going back to Engle (2002). The original proposal of Engle (2002) works well for dimensions up to the magnitude $N = 100$ and can certainly be used for the estimation of $\Sigma_{f,t}$, since the dimension of f_t is small and rarely larger than five. The problem is the estimation of $\Sigma_{u,t}$ when the number of assets is large, say $N = 1000$, which is quite common for many portfolio managers. Until recently, there did not exist a multivariate GARCH model that could accurately estimate time-varying conditional covariance matrices of such large dimensions.

The recent proposal of Engle et al. (2019) manages to do just that by combining two key innovations: first, it uses the composite-likelihood method of Pakel et al. (2017) which makes estimation in large dimensions feasible; second, it uses the nonlinear (NL) shrinkage method of Ledoit and Wolf (2012, 2015) for the estimation of the correlation targeting matrix of the DCC model, which makes the estimated matrix well-conditioned in large dimensions. We, therefore, propose the resulting DCC-NL model for the estimation of $\Sigma_{u,t}$. In this way, the estimators (2.6)–(2.8) are feasible up to dimension $N = 1000$ at least.

3.3 Averaged Forecasting of Dynamic Covariance Matrices

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, we use an 'averaged-forecasting' approach for all 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 forecast' to construct the portfolio at date h.

For the dynamics of the univariate volatilities, we use a GARCH(1,1) process:

$$d_{i,t}^2 = \omega_i + \delta_{1,i} x_{i,t-1}^2 + \delta_{2,i} d_{i,t-1}^2 \ , \tag{3.1}$$

where $(\omega_i, \delta_{1,i}, \delta_{2,i})$ 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):

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

where (δ_1, δ_2) are the DCC-NL parameters analogous to $(\delta_{1,i}, \delta_{2,i})$. 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

$$R_t := \mathsf{Diag}(Q_t)^{-1/2} \, Q_t \, \mathsf{Diag}(Q_t)^{-1/2} \quad \text{and} \tag{3.3}$$

$$\Sigma_t := D_t R_t D_t \ , \tag{3.4}$$

and the data-generating process is driven by the multivariate normal law

$$x_t | \mathcal{F}_{t-1} \sim \mathcal{N}(0, \Sigma_t) \ . \tag{3.5}$$

Hence, to determine the average of the L forecasts of the conditional covariance matrices $\Sigma_{h+l} = D_{h+l} R_{h+l} D_{h+l}$, for $l = 0, 1, \ldots, L-1$, we suggest a three-step approach where D_{h+l} and R_{h+l} are forecasted separately.

3.3.1 Step One: Forecasting Conditional Univariate Volatilities

According to Baillie and Bollerslev (1992), the multi-step ahead forecasts of the $i = 1, \ldots, N$ GARCH(1,1) volatilities can be written as

$$\mathbb{E}[d_{i,h+l}^2|\mathcal{F}_{h-1}] = \sum_{j=0}^{l-1} \omega_i(\delta_{1,i} + \delta_{2,i})^j + (\delta_{1,i} + \delta_{2,i})^l \mathbb{E}[d_{i,h}^2|\mathcal{F}_{h-1}] \ , \tag{3.6}$$

where $\mathbb{E}[d_{i,h}^2|\mathcal{F}_{h-1}] = \omega_i + \delta_{1,i}x_{i,h-1}^2 + \delta_{2,i}d_{i,h-1}^2$. Therefore, we compute the forecasts of the N-dimensional diagonal matrix D_{h+l} as

$$\mathbb{E}[D_{h+l}|\mathcal{F}_{h-1}] = \mathsf{Diag}\Big(\sqrt{\mathbb{E}[d_{1,h+l}^2|\mathcal{F}_{h-1}]}, \ldots, \sqrt{\mathbb{E}[d_{N,h+l}^2|\mathcal{F}_{h-1}]}\Big) \ . \tag{3.7}$$

3.3.2 Step Two: Forecasting Conditional Correlation Matrices

For the multivariate case we consider the approach of Engle and Sheppard (2001) where the multi-step ahead forecasts of the dynamic conditional correlation matrices are computed as

$$\mathbb{E}[R_{h+l}|\mathcal{F}_{h-1}] = \sum_{j=0}^{l-1}(1 - \delta_1 - \delta_2)C(\delta_1 + \delta_2)^j + (\delta_1 + \delta_2)^l \mathbb{E}[R_h|\mathcal{F}_{h-1}] \ , \tag{3.8}$$

using the approximation $\mathbb{E}[R_h|\mathcal{F}_{h-1}] \approx \mathbb{E}[Q_h|\mathcal{F}_{h-1}]$. In practice, the diagonal elements of the matrix C tend to deviate from one slightly, in spite of the fact that devolatized returns are used as inputs. Therefore, every column and every row has to be divided by the square root of the corresponding diagonal entry, so as to produce a proper correlation matrix.

3.3.3 Step Three: Averaging Forecasted Conditional Covariance Matrices

By using the notations $\hat{\Sigma}_{h+l} := \mathbb{E}[\Sigma_{h+l}|\mathcal{F}_{h-1}]$, $\hat{R}_{h+l} := \mathbb{E}[R_{h+l}|\mathcal{F}_{h-1}]$ and $\hat{D}_{h+l} := \mathbb{E}[D_{h+l}|\mathcal{F}_{h-1}]$ we finally calculate $\hat{\Sigma}_{h+l} := \hat{D}_{h+l}\hat{R}_{h+l}\hat{D}_{h+l}$, for $l = 0, 1, \ldots, L-1$. Therefore, to get the estimated covariance matrix on portfolio construction day h we average over the L forecasts:

$$\hat{\Sigma}_h := \frac{1}{L}\sum_{l=0}^{L-1} \hat{\Sigma}_{h+l} \ . \tag{3.9}$$

Note that in practice, the GARCH parameters in step one and the DCC-NL parameters in step two need to be estimated first. Thus, the feasible forecasts are based on $(\hat{w}_i, \hat{\delta}_{1,i}, \hat{\delta}_{2,i})$ for Equation (3.7) and on $(\hat{C}, \hat{\delta}_1, \hat{\delta}_2)$ for Equation (3.8), respectively, where $\hat{C}$ is given by the nonlinear shrinkage estimator of the unconditional correlation matrix C; see Engle et al. (2019, Section 3). This feasible averaged-forecasting scheme is then used separately for $\hat{\Sigma}_{f,h}$ and $\hat{\Sigma}_{u,h}$, as far as these inputs are needed in formulas (2.6)–(2.8).

Remark 3.1. One might ask why not use monthly data instead of daily data for the estimation of the various models given that the investment horizon in our empirical analysis of Section 4 is one month? The justification is that at the monthly frequency we do not have enough data to estimate a multivariate GARCH model. Another justification is that using daily data for the estimation seems to lead to better results even if the investment period is one month; for examples compare Tables 1 and 10 of Ledoit and Wolf (2017).[1] Therefore, if anything, we would rather go to higher frequencies towards Realized GARCH than to lower frequencies and monthly data; see Section 5. ■

3.4 Maximum Number of Observed Factors

In our context, factors are used to explain second moments. As there is a little consensus in the literature about the nature and the number of such observed 'risk factors' to be used, beyond the market factor, we feel justified in employing the five-factor model of Fama and French (2015) as the largest model in our empirical study.

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. (2017). However, factors that are successful in this context are not necessarily successful also for explaining second moments.

4 Empirical Analysis

4.1 Data and General Portfolio-Construction Rules

We download daily stock return data from the Center for Research in Security Prices (CRSP) starting on 01/01/1973 and ending on 12/31/2017. We restrict attention to stocks from the NYSE, AMEX, and NASDAQ stock exchanges. We also download daily returns on the five factors of Fama and French (2015) during the same period from the website of Ken French.

For simplicity, we adopt the common convention that 21 consecutive trading days constitute one 'month'. The out-of-sample period ranges from 01/16/1978 through 12/31/2017, resulting in a total of 480 months (or 10,080 days). All portfolios are updated monthly.[2] We denote the investment dates by $h = 1, \dots, 480$. 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

[1] Note that the two out-of-sample investment periods are not the same. Nevertheless, for $N = 1000$, using daily data for the estimation reduces the out-of-sample standard deviation of the estimated global mininum variance portfolio by 49% when upgrading from the equal-weighted portfolio to nonlinear shrinkage; on the other hand, the corresponding improvement is only 36% when using monthly data for the estimation instead.

[2] 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.

(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.[3] 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.[4] 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 have to 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.

4.2 Competing Covariance Matrix Estimators

We now detail the various covariance matrix estimators included in our empirical analysis. It is clearly of interest to also include some estimators from the literature that are not based on a factor model in order to see how their performance stacks up in comparison.

The first two estimators in our following list are, therefore, structure-free whereas the remaining ones are based on a factor model.

- **NL:** the nonlinear shrinkage estimator of Ledoit and Wolf (2017). This is a static estimator.
- **DCC-NL:** the multivariate GARCH estimator of Engle et al. (2019). This is a dynamic estimator.
- **POET:** an estimator based on an approximate factor model. This is the estimator proposed by Fan et al. (2013). The factors are sample-based and taken to be principal components of the $\{r_t\}$. In formula (2.3), $\hat{\Sigma}_f$ is given by the sample covariance matrix of the principal components used and $\hat{\Sigma}_u$ is a sparse matrix that is obtained by applying thresholding to $\mathsf{Diag}(S_{\hat{u}})$, where $S_{\hat{u}}$ is the sample covariance matrix of the $\{\hat{u}_t\}$. We use Matlab code graciously provided by the authors and keep all the default model parameters

[3]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.

[4]The reason is that we do not want to include highly similar stocks. In the early years, there are no such pairs; in the most recent years, there are never more than three such pairs.

specified in this code, which results in soft-thresholding of the residual covariance matrix. The number of principal components to be used is set to five.[5] This is a static estimator.

- **EFM:** an estimator based on an exact factor model. In formula (2.3), $\hat{\Sigma}_f$ is given by the sample covariance matrix of the $\{f_t\}$ and $\hat{\Sigma}_u$ is given by $\mathsf{Diag}(S_{\hat{u}})$. We consider both a one-factor model based on the first Fama-French factor and a five-factor model based on all five Fama-French factors. This is a static estimator.
- **AFM-POET:** an estimator based on an approximate factor model. It is similar to EFM except that $\hat{\Sigma}_u$ is obtained by applying thresholding to $S_{\hat{u}}$; in particular, the thresholding scheme is to apply the POET method to the $\{\hat{u}\}$ where the number of principal components to be used is set to zero. This is a static estimator.
- **AFM-NL:** an estimator based on an approximate factor model. It is similar to AFM-POET except that $\hat{\Sigma}_u$ is obtained by applying the nonlinear shrinkage estimator of Ledoit and Wolf (2017) to the $\{\hat{u}_t\}$. This is a static estimator.
- **AFM-DCC-NL:** an estimator based on an approximate factor model. It is similar to AFM-NL except that $\hat{\Sigma}_u$ is obtained by applying the DCC-NL estimator of Engle et al. (2019) to the $\{\hat{u}_t\}$. This is a dynamic estimator.

Note that for EFM and AFM we consider both a one-factor model based on the first Fama-French factor and a five-factor model based on all five Fama-French factors. Thereby, the number after EFM and AFM stands for the number of used factors; for example, EFM1 is an exact one-factor model and AFM5-DCC-NL is an approximate five-factor model based on DCC-NL for the covariance matrix of the error terms.

Remark 4.1. The dynamic estimator AFM-DCC-NL is based on formula (2.7) rather than on formula (2.6), that is, the dynamic part is solely due to the error terms and not also to the factors. We have chosen to follow this approach in order to facilitate the comparison with the static estimators AFM-POET and AFM-NL. Should AFM-DCC-NL perform better than these two static estimators, then we know that it has to be due to the improved estimation of the covariance matrix of the error terms (by allowing this matrix to be time-varying), since all three estimators share the same estimated covariance matrix of the factors.

Nevertheless, it is clearly of interest to study whether the performance of the estimator estimator AFM-DCC-NL can be further improved by basing it on formula (2.6) rather than on formula (2.7). Unreported numerical results reveal that the two formulas give very similar results, with neither one dominating the other. In order to save space, we therefore only report the results for formula (2.7). ■

[5] To quote from the Matlab code: "POET is very robust to overestimating K [that is, the number of factors]. But underestimating K can result in VERY BAD performance.

4.3 Global Minimum Variance Portfolio

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

$$\widehat{w} := \frac{\widehat{\Sigma}_{r,t}^{-1}\mathbb{1}}{\mathbb{1}'\widehat{\Sigma}_{r,t}^{-1}\mathbb{1}} \ . \tag{4.4}$$

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 an addition 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 $\boldsymbol{1/N}$.

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 10,080 out-of-sample returns and then multiply by 252 to annualize.
- **SD:** We compute the standard deviation of the 10,080 out-of-sample returns and then multiply by $\sqrt{252}$ to annualize.
- **IR:** We compute the (annualized) information ratio as the ratio AV/SD.

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 GMV portfolio 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 devation than another estimation model. Since we compare 12 estimation models — the exact factor model with one and five factors, four approximate factor models with one and five factors, two structure-free models, and the $1/N$ portfolio — there are 66 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 for $\hat{\Sigma}_u$, and also improves upon classical structure-free DCC-NL, we restrict attention to the comparison between the two portfolios DCC-NL and AFM-DCC-NL for one and five factors.[6] For a given universe 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).[7]

The results 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.

- With the exception of AFM-POET, all models consistently outperform $1/N$ by a wide margin.
- With the exception of AFM-POET, the approximate factor models consistently outperform the exact factor models.
- DCC-NL consistently outperforms the other structure-free models and the exact factor models. Additionally, DCC-NL outperforms in the most scenarios the approximate factor models POET, AFM-POET and AFM-NL.
- The AFM-DCC-NL model consistently outperforms the other approximate and exact factor models and for large portfolio sizes $N = 500, 1000$, we have the following overall ranking: AFM-DCC-NL, DCC-NL, AFM-NL, NL, POET, EFM, $1/N$, and AFM-POET.
- The one-factor AFM-DCC-NL consistently outperforms DCC-NL across all portfolio sizes. For large portfolio sizes $N = 500, 1000$, the outperformance is always statistically significant and also economically meaningful. The five-factor AFM-DCC-NL outperforms DCC-NL for large portfolio sizes. The outperformance is statistically significant and also economically meaningful for $N = 1000$.

[6] In Table 1 we see that DCC-NL has consistently lower SD than exact and approximate factor models. Thus, it is 'sufficient' to compare only DCC-NL and AFM-DCC-NL.

[7] Since the out-of-sample size is very large at 10,080, 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.

To sum up, dynamic estimators such as AFM-DCC-NL and DCC-NL consistently outperform static estimators.

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 (that is, with Markowitz portfolios that estimate input parameters). It can be seen that, except for AFM-POET, 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. For $N = 100$, AFM-POET is best overall, whereas for $N = 500, 1000$, AFM-DCC-NL is best overall.

Additionally, we report results on average turnover and leverage, defined as follows.

- **TO:** We compute average (monthly) turnover as $\frac{1}{479}\sum_{h=1}^{479} ||\hat{w}_{h+1} - \hat{w}_h^{\text{hold}}||_1$, where $||\cdot||_1$ denotes the L^1 norm and $\hat{w}_h^{\text{hold}}$ denotes the vector of the 'hold' portfolio weights at the end of month h.[8]
- **GL:** We compute average (monthly) gross leverage as $\frac{1}{480}\sum_{h=1}^{480} ||\hat{w}_h||_1$.
- **PL:** We compute average (monthly) proportion of leverage as $\frac{1}{480\times N}\sum_{h=1}^{480}\sum_{i=1}^{N} \mathbb{1}_{\{\hat{w}_{i,h}<0\}}$, where $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function.

The results are presented in Table 2 and can be summarized as follows; unless stated otherwise, the findings are with respect to the average monthly turnover as performance measure. Note that we do not constrain for the amounts of leverage and turnover in our optimization.

- AFM-DCC-NL consistently and markedly outperforms the structure-free DCC-NL.
- With the exception of (AFM-)POET, the dynamic models are consistently outperformed by the static models, where again the exact factor models outperform the approximate factor models.
- The amount of leverage is similar across the various models. Nevertheless, AFM-DCC-NL also consistently outperforms the structure-free DCC-NL in terms of proportion of leverage.

4.4 Markowitz Portfolio with Momentum Signal

We now turn attention to a 'full' Markowitz portfolio with a signal.

By now a large number of variables have been documented that can be used to construct a signal in practice. For simplicity and reproducibility, we use the well-known momentum factor (or simply momentum for short) of Jegadeesh and Titman (1993). For a given investment period h and a given stock, the momentum is the geometric average of the previous 252 returns on the stock but excluding the most recent 21 returns; in other words, one uses the geometric average over the previous 'year' but excluding the previous 'month'. Collecting the individual

[8]The vector $\hat{w}_h^{\text{hold}}$ is determined by the initial vector of portfolio weights, $\hat{w}_h$, together with the evolution of the prices of the N stocks in the portfolio during month h.

momentums of all the N stocks contained in the portfolio universe yields the return-predictive signal, denoted by m.

In the absence of short-sales constraints, the investment problem is formulated as

$$\min_{w} w'\Sigma_{r,t} w \tag{4.5}$$

$$\text{subject to} \quad w'm_t = b \quad \text{and} \tag{4.6}$$

$$w'\mathbb{1} = 1 \ , \tag{4.7}$$

where b is a selected target expected return. The problem has the analytical solution

$$w = c_1 \Sigma_{r,t}^{-1} \mathbb{1} + c_2 \Sigma_{r,t}^{-1} m_t \ , \tag{4.8}$$

$$\text{where } c_1 := \frac{C - bB}{AC - B^2} \quad \text{and} \quad c_2 := \frac{bA - B}{AC - B^2} \ , \tag{4.9}$$

$$\text{with } A := \mathbb{1}'\Sigma_{r,t}^{-1}\mathbb{1} \ , \quad B := \mathbb{1}'\Sigma_{r,t}^{-1} b \ , \quad \text{and} \quad C := m_t'\Sigma_{r,t}^{-1} m_t \ . \tag{4.10}$$

The natural strategy in practice is to replace the unknown $\Sigma_{r,t}$ by an estimator $\widehat{\Sigma}_{r,t}$ in formulas (4.8)–(4.10), yielding a feasible portfolio

$$\widehat{w} := c_1 \widehat{\Sigma}_{r,t}^{-1} \mathbb{1} + c_2 \widehat{\Sigma}_{r,t}^{-1} m_t \ , \tag{4.11}$$

$$\text{where } c_1 := \frac{C - bB}{AC - B^2} \quad \text{and} \quad c_2 := \frac{bA - B}{AC - B^2} \ , \tag{4.12}$$

$$\text{with } A := \mathbb{1}'\widehat{\Sigma}_{r,t}^{-1}\mathbb{1} \ , \quad B := \mathbb{1}'\widehat{\Sigma}_{r,t}^{-1} b \ , \quad \text{and} \quad C := m_t'\widehat{\Sigma}_{r,t}^{-1} m_t \ . \tag{4.13}$$

In addition to Markowitz portfolios based on formulas (4.11)–(4.13), we also include as a simple-minded benchmark the equal-weighted portfolio among the top-quintile stocks (according to momentum). This portfolio is obtained by sorting the stocks, from lowest to highest, according to their momentum and then putting equal weight on all the stocks in the top 20%, that is, in the top quintile. We denote this portfolio by **EW-TQ**. We then use the value of b implied for the EW-TQ portfolio as input in formulas (4.11)–(4.13).

Our stance is that in the context of a 'full' Markowitz portfolio, the most important performance measure is the out-of-sample information ratio, IR. In the 'ideal' investment problem (4.8)–(4.10), minimizing the variance (for a fixed target expected return b) is equivalent to maximizing the information ratio (for a fixed target expected return b). In practice, because of estimation error in the signal, the various strategies do not have the same expected return and, thus, focusing on the out-of-sample standard deviation is inappropriate.

We also consider the question whether AFM-DCC-NL delivers a higher out-of-sample information ratio than DCC-NL at a level that is statistically significant with the same reason as discussed in Section 4.3. For a given universe size, a two-sided p-value for the null hypothesis of equal information ratios is obtained by the prewhitened HAC_{PW} method described in Ledoit and Wolf (2008, Section 3.1).[9]

[9] Since the out-of-sample size is very large at 10,080, there is no need to use the computationally more expensive bootstrap method described in Ledoit and Wolf (2008, Section 3.2), which is preferred for small sample sizes.

The results are presented in Table 3 and can be summarized as follows; unless stated otherwise, the findings are with respect to the out-of-sample information ratio as performance measure.

- With the exception of AFM-POET, all models consistently outperform EW-TQ by a wide margin.
- With the exception of AFM-POET, the approximate factor models consistently outperform the exact factor models.
- DCC-NL consistently outperforms the other structure-free models and the exact factor models. Additionally, DCC-NL outperforms in the most scenarios the approximate factor models POET, AFM-POET and AFM-NL.
- With the exception of AFM-POET, the AFM-DCC-NL model consistently outperforms the other approximate and exact factor models and for large portfolio sizes $N = 500, 1000$, we have the following overall ranking: AFM-DCC-NL, DCC-NL, AFM-NL, NL, POET, EFM, EW-TQ, AFM-POET.
- The AFM-DCC-NL consistently outperforms DCC-NL for large portfolio sizes $N = 500, 1000$. For the one-factor AFM-DCC-NL with $N = 1000$ the outperformance is statistically significant and also economically meaningful.

To sum up, with the exception of AFM-POET for $N = 100$, dynamic estimators such as AFM-DCC-NL and DCC-NL consistently outperform (sophisticated) static estimators.

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 (that is, with Markowitz portfolios that estimate input parameters). Comparing with Table 1, it can be seen that, except for AFM-POET, all models based on the (simple-minded) momentum signal 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. For $N = 100$, AFM-POET is best overall, whereas for $N = 500, 1000$, AFM-DCC-NL is best overall. Even though momentum is not a very powerful return-predictive signal, the differences compared to $1/N$ can be enormous. For example, for $N = 1000$, the information ratio of $1/N$ is only 0.85 whereas the information ratio of AFM1-DCC-NL is 2.01, more than twice as large.

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. Also for this alternative performance measure, AFM-DCC-NL performs the best, followed by DCC-NL. Similar to the GMV analysis, the performance of AFM-POET is disappointing with exploding SD and thus low IR for large portfolio sizes $N = 500, 1000$.

Additionally, we report results on average turnover and leverage. The results are presented in Table 4 and can be summarized as follows; unless stated otherwise, the findings are with respect to the average monthly turnover as performance measure. Note that we do not constrain for the amounts of leverage and turnover in our optimization.

- AFM-DCC-NL consistently and markedly outperforms the structure-free DCC-NL.
- With the exception of (AFM-)POET, the dynamic models are consistently outperformed by the static models, where again the exact factor models outperform the approximate factor models.
- The amount of gross leverage is similar across the various models, apart from AFM1-POET which 'explodes' for $N = 500, 1000$ and AFM5-POET which explodes for $N = 1000$. On the other hand, AFM-DCC-NL consistently outperforms the structure-free DCC-NL in terms of proportion of leverage, although the differences are not large.

Remark 4.2. We do not provide tables with performance measures net of transaction costs for two reasons. First, we do not impose constraints on turnover, leverage, or transactions costs in any of our portfolio optimization. Of course, such constraints would be used, to varying degrees, by real-life portfolio managers; but the main point of our paper is to study the accuracy of various estimators of the covariance matrix, a problem that does not depend on transaction costs. Second, there is always disagreement which transaction cost to use. Many finance papers, at least in the past, have used a transaction cost of 50 bps. But in this day and age, the average transaction cost is usually south of 5 bps for managers trading the 1000 most liquid US stocks.

At any rate, given the results presented in Tables 1–4, the reader can get a rough idea of the various performance measures net of transaction costs, for any choice of transaction cost, according to the rule of thumb that the return loss (per month) due to turnover is twice the amount of turnover times the chosen transaction cost. For example, assuming a transaction cost of 5 bps, a turnover of one would result in a return loss of 10 bps (per month) according to this rule. ■

4.5 Summary of Results

We have carried out an extensive backtest analysis, evaluating the out-of-sample performance of our dynamic approximate factor model based on a new DCC-NL estimation scheme. Consequently, we have compared AFM-DCC-NL to a number of other strategies — various factor models and structure-free estimators of the covariance matrix — to estimate the global mininum-variance portfolio and the Markowitz portfolio with momentum signal. Among the considered portfolios, AFM-DCC-NL is the clear winner. In most scenarios, AFM-DCC-NL performs the best, followed by DCC-NL. Only for a small investment universe $N = 100$, the dynamic models are outperformed by AFM-POET for the Markowitz portfolios with momentum signal. In all other scenarios the dynamic estimators consistently outperform the static estimators. Note that for large investment universes the long and short portfolio holdings of AFM-POET explode, which returns very high (low) out-of-sample SD (IR). Additionally, for large portfolio sizes the outperformance of AFM-DCC-NL over DCC-NL is statistically significant and economically meaningful. Thus, imposing some structure via an approximate factor model improves the out-of-sample performance.

However, including multiple factors (that is, using 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 time-varying conditional heteroskedasticity; on the other hand, additional factors do not seem to be needed for sophisticated estimation methods.

In sum, we recommend the dynamic approximate one-factor model based on DCC-NL, which outperforms its five-factor 'cousin'. This finding makes the resulting AFM1-DCC-NL estimator even more attractive for industry and applied portfolio management, since only data on the market factor are needed. Therefore, this estimator can be easily implemented also outside of the US, in countries for which all five Fama-French factors are not readily available.

5 Outlook on Future Research

In this paper, we have used daily data for the estimation of the various models. Although this is, arguably, still the most common approach both in the literature and in the industry, there have been several recent proposals on using intra-daily data instead for the estimation of models used for portfolio selection and asset pricing based on realized measures of daily covariance matrices; for example, see Bauer and Vorkink (2011), Chiriac and Voev (2011), Lunde et al. (2016), Callot et al. (2017), and Aït-Sahalia and Xiu (2017). Most of these papers only consider a limited number of assets, typically less than 50, but Brito et al. (2018) extend the work of Callot et al. (2017) by proposing a Factor-HAR model estimated with penalized regressions to forecast large realized covariance models of dimension $N = 430$. On the other hand, it is not clear whether their proposal could also handle dimensions of $N = 1000$ and beyond, as trading asynchronicity poses a problem in the construction of realized measures of daily covariances and this problem (greatly) increases with the dimension N.

It would, therefore, be of interest to combine our proposal with this strand of literature. An obvious first idea is to use the Realized GARCH model of Hansen et al. (2012), which uses intra-daily data, instead of the GARCH(1,1) model (3.1) in the first step of the DCC-NL model. Doing so should yield more accurate forecasts of the conditional univariate volatilities and thus lead to improved performance of the DCC-NL model, whether used by itself or as in input in the AFM-DCC-NL model. In this way, intra-daily data would 'only' be used to forecast the conditional univariate volatilities, whereas the forecasts of the conditional correlations would still use daily data according to (3.2). The advantage of this 'hybrid' approach is that it does not introduce any limitation due to intra-daily data as far as the dimension N is concerned: One can handle the same dimensions that can be handled by our proposed models in this paper, which only use daily data.

A second idea is to incorporate shrinkage in the forecasting of the residual covariance matrix $\Sigma_{u,t}$ in the methodology of Brito et al. (2018). To counter the curse of dimensionality, Brito et al. (2018, Section 3.3) impose a block-diagonal structure for $\Sigma_{u,t}$, where the blocks correspond to industries. An alternative would be not to impose any structure and to post-process the forecasted matrix $\Sigma_{u,t}$ with shrinkage instead in order to obtain a well-conditioned matrix. In particular, such an approach is able to capture extra-factor covariance between industries.

6 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, especially when allowing for time-varying conditional heteroskedasticity, as in our new AFM1-DCC-NL model. As a secondary contribution, we propose a novel scheme for extrapolating the covariance matrix forecast over the holding period of the investment strategy in case the holding period exceeds the frequency of the observed data (such as when the holding period is a month and the observed data are daily). 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 along the lines suggested by Ledoit et al. (2019).

References

Aït-Sahalia, Y. and Xiu, D. (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics*, 201(2):384–399.

Ang, A. and Kristensen, D. (2012). Testing conditional factor models. *Journal of Financial Economics*, 106:132–156.

Avramov, D. and Chordia, T. (2006). Asset pricing models and financial market anomalies. *Review of Financial Studies*, 19(3):1001–1040.

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

Bai, J. and Shi, S. (2011). Estimating high dimensional covariance matrices and its applications. *Annals of Economics and Finance*, 12(2):199–215.

Baillie, R. T. and Bollerslev, T. (1992). Prediction in dynamic models with time-dependent conditional variances. *Journal of Econometrics*, 52:91–113.

Bali, T. G., Engle, R. F., and Tang, Y. (2017). Dynamic conditional beta is alive and well in the cross section of daily stock returns. *Management Science*, 63(11):3760–3779.

Bauer, G. H. and Vorkink, K. (2011). Forecasting multivariate realized stock market volatility. *Journal of Econometrics*, 160(1):93–101.

Brito, D., Medeiros, M. C., and Ribeiro, R. (2018). Forecasting large realized covariance matrices: The benefits of factor models and shrinkage. Technical report. Available at `https://ssrn.com/abstract=3163668`.

Callot, L. A., Kock, A. B., and Medeiros, M. C. (2017). Modeling and forecasting large realized covariance matrices and portfolio choice. *Journal of Applied Econometrics*, 32(1):140–158.

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

Chiriac, R. and Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 26(6):922–947.

Connor, G. and Korajczyk, R. A. (1993). A test for the number of factors in an approximate factor model. *Journal of Finance*, 48:1263–1291.

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

DeMiguel, V., Garlappi, L., and Uppal, R. (2009b). Optimal versus naive diversification: How inefficient is the $1/N$ portfolio strategy? *Review of Financial Studies*, 22:1915–1953.

DeMiguel, V., Martin-Utrera, A., and Nogales, F. J. (2013). Size matters: Optimal calibration of shrinkage estimators for for portfolio selection. *Journal of Banking & Finance*, 37:3018–3034.

Engle, R. F. (2002). Dynamic conditional correlation — a simple class of multivariate GARCH models. *Journal of Business & Economics Statistics*, 20:339–350.

Engle, R. F. (2016). Dynamic conditional beta. *Journal of Financial Econometrics*, 14(4):643–667.

Engle, R. F. and Colacito, R. (2006). Testing and valuing dynamic correlations for asset allocation. *Journal of Business & Economic Statistics*, 24(2):238–253.

Engle, R. F., Ledoit, O., and Wolf, M. (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*. Forthcoming.

Engle, R. F. and Sheppard, K. (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Working Paper 8554, National Bureau of Economic Research.

Fama, E. F. and French, K. R. (2015). A five factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

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

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements (with discussion). *Journal of the Royal Statistical Society, Series B*, 75(4):603–680.

Feng, G., Giglio, S., and Xiu, D. (2017). Taming the factor zoo. Research Paper 17-04, Chicago Booth. Available at `https://ssrn.com/abstract=2934020`.

Green, J., Hand, J. R. M., and Zhang, X. F. (2013). The supraview of return predictive signals. *Review of Accounting Studies*, 18:692–730.

Hansen, P. R., Huang, Z., and Shek, H. H. (2012). Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906.

Haugen, R. A. and Baker, N. L. (1991). The efficient market inefficiency of capitalization-weighted stock portfolios. *Journal of Portfolio Management*, 17(3):35–40.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 54(4):1651–1684.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1):65–91.

Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10(5):603–621.

Ledoit, O. and Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, 15:850–859.

Ledoit, O. and Wolf, M. (2011). Robust performance hypothesis testing with the variance. *Wilmott Magazine*, September:86–89.

Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics*, 40(2):1024–1060.

Ledoit, O. and Wolf, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. *Journal of Multivariate Analysis*, 139(2):360–384.

Ledoit, O. and Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *Review of Financial Studies*, 30(12):4349–4388.

Ledoit, O., Wolf, M., and Zhao, Z. (2019). Efficiet sorting: A more powerful test for cross-sectional anomalies. *Journal of Financial Econometrics*. Forthcoming.

Lunde, A., Shephard, N., and Sheppard, K. (2016). Econometric analysis of vast covariance matrices using composite realized kernels and their application to portfolio choice. *Journal of Business & Economic Statistics*, 34(4):504–518.

Meucci, A. (2005). *Risk and Asset Allocation*. Springer Finance, Berlin Heidelberg New York.

Nielsen, F. and Aylursubramanian, R. (2008). Far from the madding crowd — Volatility efficient indices. Research insights, MSCI Barra.

Pakel, C., Shephard, N., Sheppard, K., and Engle, R. F. (2017). Fitting vast dimensional time-varying covariance models. Working Paper FIN-08-009, NYU. Available at `https://ssrn.com/abstract=1354497`.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360.

Stock, J. H. and Watson, M. (2011). Dynamic factor models. In Clements, M. P. and Hendry, D. F., editors, *The Oxford Handbook on Economic Forecasting*, pages 35–60. Oxford University Press, Oxford.

A Tables

Period: 01/16/1978–12/31/2017									
	$N = 100$			$N = 500$			$N = 1000$		
	AV	SD	IR	AV	SD	IR	AV	SD	IR
Structure-Free Models									
$1/N$	12.82	17.40	0.74	13.86	16.83	0.82	14.36	16.85	0.85
NL	11.94	11.74	1.02	11.91	8.63	1.38	12.28	7.45	1.65
DCC-NL	11.62	11.59	1.00	12.57	8.26	1.52	12.84	6.93	1.85
Exact Factor Models									
EFM1	13.06	14.12	0.93	12.52	12.14	1.03	13.35	10.97	1.22
EFM5	13.02	12.68	1.03	12.68	10.97	1.16	12.90	9.72	1.33
Approximate Factor Models									
POET	12.04	11.98	1.00	11.86	8.48	1.40	13.09	7.82	1.67
AFM1-POET	12.40	11.75	1.06	15.87	74.07	0.21	−763	3962	−0.19
AFM5-POET	12.25	11.71	1.05	12.19	8.72	1.40	−27	295	−0.09
AFM1-NL	11.97	11.75	1.02	11.90	8.63	1.38	12.28	7.45	1.65
AFM5-NL	11.95	11.76	1.02	11.88	8.63	1.38	12.20	7.45	1.64
AFM1-DCC-NL	11.55	**11.56**	1.00	12.65	**8.11***	1.56	13.31	**6.61*****	2.01
AFM5-DCC-NL	11.53	11.64	0.99	12.53	8.18	1.53	12.92	6.65***	1.94

Table 1: Annualized performance measures (in percent) for various estimators of the GMV portfolio. AV stands for average; SD stands for standard deviation; and IR stands for information ratio. The number after EFM and AFM stands for the number of considered Fama-French factors. All measures are based on 10,080 daily out-of-sample returns from 01/16/1978 until 12/31/2017. In the columns labeled SD, the lowest number appears in **bold face**. In the rows labeled AFM1-DCC-NL and AFM5-DCC-NL, significant outperformance over DCC-NL 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.

Period: 01/16/1978–12/31/2017									
	$N = 100$			$N = 500$			$N = 1000$		
	TO	GL	PL	TO	GL	PL	TO	GL	PL
No Factor Models									
$1/N$	0.11	1.00	0.00	0.10	1.00	0.00	0.10	1.00	0.00
NL	0.54	2.64	0.43	0.84	3.65	0.42	0.85	3.54	0.42
DCC-NL	1.85	2.75	0.46	2.12	3.24	0.49	1.89	2.98	0.50
Exact Factor Models									
EFM1	0.32	2.23	0.44	0.25	2.22	0.45	0.22	2.10	0.45
EFM5	0.38	2.44	0.43	0.39	2.85	0.45	0.35	2.70	0.46
Approximate Factor Models									
POET	0.53	2.64	0.43	0.70	3.47	0.46	1.69	4.22	0.47
AFM1-POET	0.53	2.66	0.43	38.21	10.11	0.46	393.18	104.75	0.48
AFM5-POET	0.52	2.68	0.42	1.10	4.19	0.46	60.91	32.64	0.48
AFM1-NL	0.55	2.69	0.41	0.86	3.68	0.43	0.87	3.56	0.42
AFM5-NL	0.55	2.71	0.41	0.86	3.71	0.43	0.89	3.60	0.42
AFM1-DCC-NL	1.04	2.70	0.42	1.45	3.45	0.46	1.36	3.42	0.47
AFM5-DCC-NL	0.94	2.76	0.42	1.29	3.62	0.45	1.25	3.50	0.46

Table 2: Average monthly turnover and leverage for various estimators of the GMV portfolio. TO stands for turnover; GL stands for gross leverage; and PL stands for proportion of leverage. The number after EFM and AFM stands for the number of considered Fama-French factors. All measures are averages based on 480 monthly portfolio weight vectors from the out-of-sample period 01/16/1978 until 12/31/2017.

Period: 01/16/1978–12/31/2017									
	$N = 100$			$N = 500$			$N = 1000$		
	AV	SD	IR	AV	SD	IR	AV	SD	IR
Structure-Free Models									
EW-TQ	16.55	21.33	0.78	16.85	20.24	0.83	17.55	20.30	0.87
NL	14.76	14.16	1.04	14.54	10.10	1.44	15.00	8.75	1.71
DCC-NL	14.95	14.13	1.06	14.87	9.51	1.56	14.82	7.95	1.86
Exact Factor Models									
EFM1	15.37	16.50	0.93	15.52	13.93	1.11	16.33	12.78	1.28
EFM5	15.22	15.49	0.98	15.76	12.80	1.23	15.94	11.39	1.40
Approximate Factor Models									
POET	14.53	14.33	1.01	14.28	10.02	1.43	15.45	9.10	1.70
AFM1-POET	15.14	14.17	**1.07**	−82	438	−0.19	338	3592	0.09
AFM5-POET	14.89	14.17	1.05	14.69	10.20	1.44	−70	327	−0.21
AFM1-NL	14.79	14.16	1.04	14.52	10.09	1.38	15.00	8.75	1.72
AFM5-NL	14.78	14.17	1.04	14.48	10.10	1.44	14.90	8.75	1.70
AFM1-DCC-NL	14.69	14.02	1.05	15.24	9.46	**1.61**	15.76	7.84	**2.01*****
AFM5-DCC-NL	14.58	14.09	1.04	14.97	9.58	1.56	15.28	7.91	1.93

Table 3: Annualized performance measures (in percent) for various estimators of the Markowitz portfolio with momentum signal. AV stands for average; SD stands for standard deviation; and IR stands for information ratio. The number after EFM and AFM stands for the number of considered Fama-French factors. All measures are based on 10,080 daily out-of-sample returns from 01/16/1978 until 12/31/2017. In the columns labeled IR, the largest number appears in **bold face**. In the rows labeled AFM1-DCC-NL and AFM5-DCC-NL, significant outperformance over DCC-NL in terms of IR 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.

Period: 01/16/1978–12/31/2017									
	$N = 100$			$N = 500$			$N = 1000$		
	TO	GL	PL	TO	GL	PL	TO	GL	PL
No Factor Models									
EW-TQ	0.60	1.00	0.00	0.57	1.00	0.00	0.57	1.00	0.00
NL	1.12	3.12	0.42	1.32	4.07	0.43	1.28	3.95	0.42
DCC-NL	2.32	3.35	0.46	2.49	3.75	0.49	2.27	3.51	0.50
Exact Factor Models									
EFM1	0.92	2.73	0.44	0.84	2.71	0.44	0.80	2.59	0.43
EFM5	0.95	2.90	0.43	0.95	3.28	0.45	0.93	3.17	0.46
Approximate Factor Models									
POET	1.11	3.14	0.43	1.24	3.94	0.46	2.26	4.72	0.47
AFM1-POET	1.12	3.17	0.43	17.97	13.52	0.47	243.44	89.97	0.48
AFM5-POET	1.11	3.19	0.43	1.66	4.69	0.46	129.43	55.37	0.47
AFM1-NL	1.12	3.19	0.42	1.33	4.10	0.43	1.30	3.98	0.42
AFM5-NL	1.12	3.21	0.42	1.33	4.13	0.43	1.31	4.01	0.42
AFM1-DCC-NL	1.56	3.24	0.43	1.92	4.06	0.47	1.85	4.00	0.47
AFM5-DCC-NL	1.46	3.28	0.43	1.78	4.12	0.46	1.75	4.04	0.47

Table 4: Average monthly turnover and leverage for various estimators of the Markowitz portfolio with momentum signal. TO stands for turnover; GL stands for gross leverage; and PL stands for proportion of leverage. The number after EFM and AFM stands for the number of considered Fama-French factors. All measures are averages based on 480 monthly portfolio weight vectors from the out-of-sample period 01/16/1978 until 12/31/2017.