TI 2018-099/III
Tinbergen Institute Discussion Paper

Estimation Risk and Shrinkage in Vast-Dimensional Fundamental Factor Models

Andries C. van Vlodrop[1]
Andre (A.) Lucas[1]

[1] Vrije Universiteit Amsterdam, The Netherlands

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

Estimation Risk and Shrinkage in Vast-Dimensional Fundamental Factor Models*

Andries van Vlodrop[(a)] André Lucas[(a)]

[(a)] *Vrije Universiteit Amsterdam and Tinbergen Institute*

December 11, 2018

Abstract

We investigate covariance matrix estimation in vast-dimensional spaces of 1,500 up to 2,000 stocks using fundamental factor models (FFMs). FFMs are the typical benchmark in the asset management industry and depart from the usual statistical factor models and the factor models with observed factors used in the statistical and finance literature. Little is known about estimation risk in FFMs in high dimensions. We investigate whether recent linear and non-linear shrinkage methods help to reduce the estimation risk in the asset return covariance matrix. Our findings indicate that modest improvements are possible using high-dimensional shrinkage techniques. The gains, however, are not realized using standard plug-in shrinkage parameters from the literature, but require sample dependent tuning.

1 Introduction

Since the introduction of mean-variance portfolio optimization (Markowitz, 1952), estimating the covariance matrix of asset returns is one of the main

*This research was performed while van Vlodrop was at APG Asset Management. We thank Gerben de Zwart, Marco Della Seta, and Anne Opschoor for useful feedback.

challenges in portfolio management. The task can be daunting, as the dimension of this covariance matrix is typically large in empirically relevant applications. For instance, a number of $1,500$ stocks or more is not exceptional in a standard stock portfolio setting. Such large numbers of assets introduce an estimation problem: the number of correlations needed for the covariance matrix increases quadratically with the number of assets. The estimation risk may become so overwhelming that simple diversification techniques that avoid estimation altogether can perform better ex-post than ex-ante optimal investment strategies based on estimated parameters; see for instance the "naive $1/N$" strategy of DeMiguel et al. (2007).

One approach to overcoming the estimation risk problem of the covariance matrix in high dimensions is the use of factor models for asset returns. By imposing a factor model structure, the number of parameters is of the order of the number of assets times the number of factors. For a limited number of factors compared to the number of assets, this results in sizable reductions of the parameter space. Factor models come in three different blends: (1) models where factors are observed but factor loadings are not, such as seminal papers by Fama and MacBeth (1973), Ross (1976), and many more; (2) models where neither factors nor factor loadings are oberved, such as seminal papers by for instance Connor and Korajczyk (1986), Brown (1989); and (3) models where factor loadings are observed, but factors are not, such as Menchero et al. (2008). Models of type (1) and (2) have been elaborately discussed in the finance and statistics literature. Surprisingly, however, models of type (3) have received much less attention in the academic literature. These models are also known as Fundamental Factor Models (FFMs). The lack of attention to FFMs is the more surprising given that these models are the typical benchmark used in the portfolio management industry for asset allocation, risk attribution, etc.

In this paper, we fill this gap by studying the effect of estimation risk on FFMs in more detail. In particular, we are interested in whether the FFM's structure itself already suffices for overcoming estimation risk problems in high-dimensional covariance matrices, or whether additional tools are required, such as statistical shrinkage techniques. Statistical shrinkage relates to a set of techniques that tilt a standard estimator for the covariance matrix to a target based on outside or prior information. In the presence of estimation risk, such tilting typically improves upon the standard estimator.

Two popular methods advocated in the literature are linear and non-linear shrinkage. Linear shrinkage methods are applied most often and have a clear intuitive appeal. A linear shrinkage estimator for the covariance matrix is a convex combination of a standard covariance matrix estimator that is subject to estimation risk, and a well-behaved shrinkage target. The shrinkage target may take different forms, such as a covariance matrix with correlations equal to zero, correlations equal to a single value (equi-correlation), identical variances, etc. The weight assigned to the standard estimator and the shrinkage target is controlled by the shrinkage coefficient. Methods for optimally choosing the shrinkage coefficient in high dimensions have recently been developed by for instance Ledoit and Wolf (2003, 2004b). The practical usefulness of these methods is that the optimal value of the shrinkage coefficient can be estimated based on the sample of asset returns. In an application of the method, Ledoit and Wolf (2004a) use an equicorrelation target and find clear improvements over the sample covariance matrix in a portfolio optimization context.

More recently, non-linear shrinkage methods have gained popularity, see Engle et al. (2017); Lam et al. (2016); Ledoit and Wolf (2015, 2017a,b, 2018). These methods succeed in inverting the asymptotic bias in the eigenvalue-spectrum in large dimensions and long samples, i.e., so-called large (N, T)-asymptotics. Non-linear shrinkage is computationally more demanding than linear shrinkage, but comes with the advantage that one does not have to set an explicit shrinkage target. To make non-linear shrinkage more accessible for practitioners, Ledoit and Wolf (2017a) propose a computationally efficient method to implement non-linear shrinkage.

It is as yet unknown how FFMs can benefit from recent statistical shrinkage methods, if at all. Given the benchmark status of FFMs in the industry, the relevance of obtaining more insight into this questions is evident. In this paper we therefore evaluate the usefulness of shrinkage methods after imposing a factor structure on the covariance matrix. While the FFM provides a dimension reduction with respect to the sample covariance matrix in terms of the number of free parameters, the number of factors can still be relatively large compared to the length of the time series. To illustrate this, we use the set-up of the Barra GEM2 model, described in Menchero et al. (2008).[1] When including country, industry, and style factors, the total number of factors can amount to up to 60. Combined with a typical estimation window of five years of monthly data (60 observations), the challenge in terms of estimation risk is evident.

To evaluate the accuracy of different covariance matrix estimators and shrinkage techniques, we perform an out-of-sample analysis. Each month, we use a rolling estimation period of five years to estimate the asset return covariance matrix for each of the methods. We then compute the implied minimum volatility portfolio and evaluate the one-month out-of-sample performance of this portfolio for each of the methods considered.Our set-up mimics the situation of institutional investors considering optimal portfolios in a stock universe comparable in size to the MSCI World Index.

We find that the standard FFM without any shrinkage already shows a good performance. The factor structure imposed already takes care of many of the estimation risk challenges that are solved by shrinkage of the sample covariance matrix elsewhere in the literature. Linear and non-linear shrinkage techniques applied to the covariance matrix of the fundamental factor returns even result in a worse performance during most periods in the sample: the bias in the fundamental factor returns' covariance matrix induced by shrinkage offsets any potential gains in reduced estimation risk. We find that

[1]We use a simplified version of this methodology, such that our results cannot be interpreted as a statement on the performance of the Barra GEM2 model. Instead, we only use GEM2 as a typical example of a fundamental factor model to assess the performance of these models in combination with or in comparison to shrinkage methods.

the worse performance is caused by standard plug-in values for the shrinkage coefficient. The analysis further indicates that the potential empirical gains from shrinking the factor returns covariance matrix are quite limited, even if the optimal bandwidth parameter could be found. A more promising route for shrinkage in the FFM is to shrink the idiosyncrated covariance matrix to a shrinkage target based on equal idiosyncratic variances. Also for this approach, however, the gains are limited and not statistically significant. Allowing for time variation in the underlying factor return covariance matrix or the idiosyncratic covariance matrix does not alter our results. We conclude that FFMs provide a good benchmark for overcoming estimation risk by imposing structure on asset returns. Recently proposed shrinkage estimators appear to provide a good competitive alternative approach, but as yet do not outperform the typical industry benchmark.

In a recent independent study Brito et al. (2018) also investigate shrinkage in FFMs. Our set-up differs from their set-up in several important aspects. In particular, we consider a vast dimension of 1,500 stocks or more, whereas Brito et al. include a much smaller number of stocks of only 430. Our numbers of substantially over 1,000 assets are much more in line with what is needed empirically. Second, we follow the industry practice and base the factor loadings on observed company characteristics. It is precisely in this context that the question on the effectiveness of shrinkage techniques in vast-dimensional FFM covariance matrices is most relevant.

The rest of this paper is organized as follows. Section 2 discusses in detail the FFM and the ways to reduce estimation risk. In Section 3 we discuss the data and the evaluation set-up. Section 4 presents the results, and Section 5 concludes.

2 Estimating the covariance matrix

2.1 The fundamental factor risk model

Our objective is to model the $N \times N$ covariance matrix of excess stock returns for $N > 1,500$. Let R_t denote the N-dimensional vector of returns (in Euros), and let r_t^f denote a proxy for the euro area risk-free interest rate over the same period. The fundamental factor model (FFM) for the excess returns $R_t - r_t^f$ is specified as

$$R_t - r_t^f = X_t f_t + X_t^C f_t^C + u_t, \tag{1}$$

where the K-dimensional vector f_t contains the (unobserved) factor returns over period t, and X_t represents the $N \times K$ matrix of factor loadings as observed at the start of the period. The $N \times K^C$ matrix X_t^C and the $K^C \times 1$ vector f_t^C contain (dummy indicator) currency exposures and corresponding observed currency returns, respectively. Finally, the $K \times 1$ vector u_t represents the unobserved idiosyncratic part of the stock returns.

We consider the risk factors of the Barra GEM2 model of Menchero et al. (2008), i.e. $f_t' = [f_t^{1\prime}\ f_t^{2\prime}\ f_t^{3\prime}]$, where f_t^1, f_t^2, and f_t^3 represent the industry, country, and style factor returns, respectively (see Menchero et al., 2008, for more details). Note that unlike the standard factor model with observed f_t and unobserved X_t, or the statistical factor model with both f_t and X_t unobserved, the FFM has X_t as observed and f_t as unobserved, comparable to a second-pass regression of the Fama-MacBeth type (see Fama and MacBeth, 1973).

We obtain estimates of the factor returns $\hat{f}_t$ from weighted least squares estimation of

$$r_t - r_t^f = X_t f_t + u_t, \tag{2}$$

where r_t represents the local stock returns, i.e. r_t excludes the currency returns, such that $R_t = r_t + X_t^C f_t^C$ in (1).[2] We estimate the common risk

[2]We use weighted least squares estimation since the idiosyncratic stock returns are

factors $\hat{f}_t$ for each time period $t = 1, \ldots, T$, using cross-sectional regressions.[34] This provides us a time series of estimated factor returns $\hat{f}_1, \ldots, \hat{f}_T$, and corresponding idiosyncratic returns $\hat{u}_1, \ldots, \hat{u}_T$.

We assume that the idiosyncratic returns u_t are zero mean and independent of the factor returns, both cross-sectionally and over time. The expression for the covariance matrix Σ_t of the excess Euro returns then follows as

$$\Sigma_t \equiv \text{cov}(R_t - r_t^f) = [X_t \; X_t^C] \; F \; [X_t \; X_t^C]' + U, \tag{3}$$

where F represents the covariance matrix of the factor and currency returns, i.e., of the vector $[f_t{}' \; f_t^{C\prime}]'$, and the diagonal matrix U collects the idiosyncratic variances. Using the time series of estimated $\hat{f}_1, \ldots, \hat{f}_T$ and $\hat{u}_1, \ldots, \hat{u}_T$ together with the currency returns $f_1^C, \ldots, f_T^C$, we obtain the estimates $\hat{F}$ and $\hat{U}$.[5] The covariance matrices F and U may be estimated by a sample covariance matrix, an exponentially weighted moving average (EWMA) scheme, or by any other method. The estimated covariance matrix $\hat{\Sigma}_t$ of excess returns then follows by plugging $\hat{F}$ and $\hat{U}$ into (3).

2.2 Estimation risk in the covariance matrix

The estimated covariance matrix $\hat{\Sigma}_t$ based on (3) is subject to estimation risk when $\hat{F}$ and $\hat{U}$ are estimated in finite samples. In particular, it might

heteroskedastic. We follow Menchero et al. (2008) and use marketcap$^{1/4}$ for the weights. This puts more weight on stocks with a larger market capitalization, which typically are the stocks with a lower idiosyncratic volatility.

[3]For robustness purposes we truncate the local returns when performing weighted least squares estimation in (2), similar to the methodology of Menchero et al. (2008). More precisely, we calculate their standard deviation and then truncate these local returns at three standard deviations from their mean return.

[4]Since the factor loadings for the industry and country factors are dummies and sum up to one for each stock there is collinearity in X_t. To solve this issue, we follow Menchero et al. (2008) and first calculate for each industry and country the total marketcap of stocks belonging to it. Afterwards, we impose the restriction that the marketcap weighted sum of industry as well as country factor returns equals zero.

[5]Note that $\hat{F}$ estimates F using the *estimated* $\hat{f}_1, \ldots, \hat{f}_T$ rather than the (unobserved) true $f_1, \ldots, f_T$. Since our cross-sectional dimension is large, we appeal to a consistency argument and argue that $\hat{F}$ and $\hat{U}$ are still consistent estimates of F and U, respectively. Note that estimation error in $\hat{f}_t$ and $\hat{u}_t$ may have non-trivial effects on the inference.

be difficult to precisely pin down the individual correlation terms in $\hat{F}$ based on a short sample. In total there are $K^* = K + K^C$ factors including the currency returns. This yields $K^*(K^* - 1)/2$ correlations. For example, if we take a typical industry standard of $K^* = 75$ factors, the corresponding number of correlations to be estimated is $2,775$. Furthermore, the diagonal matrix $\hat{U}$ typically contains around $N = 1,500$ individual variances that also need to be estimated. Given these numbers, it is clear that in finite samples the $\hat{F}$ and $\hat{U}$ matrices are potentially subject to a considerable degree of estimation risk. This estimation risk may subsequently spill over to $\hat{\Sigma}_t$. To overcome estimation risk, we can use different shrinkage techniques, which we discuss next.

2.2.1 Linear shrinkage

A common method to deal with estimation risk is to use a linear shrinkage estimator. The concept of linear shrinkage is intuitive. Consider the sample covariance matrix $\hat{F}$ of the factor returns and a well-behaved shrinkage target $\hat{F}^0$. The target $\hat{F}^0$ is misspecified, but subject to a lower degree of estimation risk. For instance, $\hat{F}^0$ my put a number of elements equal to each other (such as the correlations), or restrict elements to particular values (such as setting the correlations to zero). The linear shrinkage estimator $\hat{F}^S$ is obtained as a convex combination of the sample covariance matrix and the target,

$$\hat{F}^S = (1 - \delta)\hat{F} + \delta\hat{F}^0, \tag{4}$$

where $\delta \in [0, 1]$ represents the weight of the shrinkage target. It is quite common for the "optimal" shrinkage estimator that $0 < \delta < 1$, i.e., we can improve on $\hat{F}$ by putting it closer to $\hat{F}^0$, see e.g. Ledoit and Wolf (2004a).

We propose several candidates for the shrinkage target $\hat{F}^0$. In our first target choice, we specify a version with a parsimonious correlation structure building on the Barra GEM2 FFM specification. In particular, we use a target that sets the correlation between classes of risk factors to zero, and uses an equicorrelation structure within a given risk factor class. We have

$$\hat{F}^0 = \begin{bmatrix} \hat{F}_1^0 & 0 & 0 & 0 \\ 0 & \hat{F}_2^0 & 0 & 0 \\ 0 & 0 & \hat{F}_3^0 & 0 \\ 0 & 0 & 0 & \hat{F}_C^0 \end{bmatrix}, \tag{5}$$

where $\hat{F}_1^0$, $\hat{F}_2^0$, $\hat{F}_3^0$, and $\hat{F}_C^0$ represent the shrinkage targets for the covariance matrices of the industry, country, style, and currency factor returns, respectively. Each of these matrices is based on an equicorrelation structure that we obtain from the sample covariance matrices in a straightforward way. For instance, if $\hat{F}_1$ is the sample covariance matrix of the industry factors, we have

$$\hat{F}_1^0 = \hat{\Delta}_1^{1/2} \left((1-\hat{\rho}_1)\mathrm{I} + \hat{\rho}_1\, \iota\, \iota' \right) \hat{\Delta}_1^{1/2},$$
$$\hat{\rho}_1 = \left(\iota' \hat{\Delta}_1^{-1/2}\, \hat{F}_1\, \hat{\Delta}_1^{-1/2} \iota - K_1 \right) \,/\, \left(K_1(K_1-1) \right),$$

where $\hat{\Delta}_1$ is a diagonal matrix containing the diagonal of $\hat{F}_1$, K_1 is the dimension of F_1, I is an identity matrix, and ι is a vector of ones. This sets the equicorrelation in $\hat{F}_1^0$ equal to the average correlation in $\hat{F}_1$.

As our second shrinkage target, we use an equicorrelation of zero, i.e., we set $\hat{\rho}_i = 0$ for $i = 1, \ldots, 4$. Together, these two shrinkage targets reduce estimation risk by shrinking the correlations within groups of factor returns towards equal values (and to zero for the second version). The correlations between factor returns from different groups are shrunk towards zero. Combined, this reduces the effect of spurious correlation values observed in the sample covariance matrix.

We also experiment with a shrinkage target for the diagonal matrix $\hat{U}$. In particular, we specify the shrinkage target $\hat{U}^0$ as a version with equal idiosyncratic variances, i.e.,

$$\hat{U}^0 = \bar{\sigma}^2 \mathrm{I}_N \tag{6}$$

where $\bar{\sigma}^2$ equals the average idiosyncratic sample variance over the N stocks.

2.2.2 Setting the linear shrinkage parameter

The shrinkage intensity δ plays an important role in the analysis. If $\delta = 0$ there is no shrinkage and also no potential gains. If $\delta = 1$ the shrinkage is dogmatic, however, the bias in the shrinkage target may be too large. In principle, δ needs to shrink to zero as the sample size diverges to infinity. This, however, provides little guidance in finite samples. To set δ, we follow the asymptotic arguments derived in Ledoit and Wolf (2003, 2004b), who minimize the expected Frobenius norm of the difference between the shrinkage estimator and the true covariance matrix F. This optimal value of δ, denoted as δ^*, is given by

$$\delta^* = \text{argmin}_{\delta \in [0,1]} \mathbb{E}\left[||(1-\delta)\hat{F} + \delta \hat{F}^0 - F||^2 \right], \tag{7}$$

where we assume that F is time-invariant. Similarly, for the idiosyncratic variances we have

$$\delta^* = \text{argmin}_{\delta \in [0,1]} \mathbb{E}\left[||(1-\delta)\hat{U} + \delta \hat{U}^0 - U||^2 \right]. \tag{8}$$

Ledoit and Wolf (2003, 2004b) show that we can consistently estimate the optimal δ^* for fixed N when $T \to \infty$, implementation details can be found in in Appendix A. Since we only require in-sample information for the estimate $\hat{\delta}^*$, this is a valid method in our out-of-sample evaluation of the different shrinkage methods. Note, however, that our estimation samples are as small as $T = 60$, whereas the number of stocks N is typically larger than 1,500. This casts doubt on the applicability of arguments for the determination of δ based on large T asymptotics. Therefore, the determination of δ is also part of our analysis later on.

2.2.3 Non-linear shrinkage

Recently important steps have been made in the literature that considers non-linear shrinkage estimators. These estimators operate directly on the eigenvalues of large covariance matrices, the so-called eigenvalue spectrum. The underlying mathematics for non-linear shrinkage estimators originates from the field of Random Matrix Theory; see the review in Ledoit and Wolf (2017a). For large N and T, with N/T tending to a constant, the eigenvalue spectrum converges to a highly specific curve, which causes a bias in correlation matrix estimates in vast dimensions. Ledoit and Wolf (2017a) provides a numerical solution to the inversion of this bias problem, which gives a way to obtain consistent estimates in large dimensions. These were successfully applied even in time-varying parameter cases such as the DCC model, see Engle et al. (2017).

Numerically, the recent non-linear shrinkage estimators are as fast as the linear shrinkage estimators. We therefore include these new techniques in our comparison. We apply the non-linear shrinkage estimator of Ledoit and Wolf (2017a) to the $\hat{F}$ matrix in two ways. First, we use the non-linear shrinkage estimator on the full $\hat{F}$ matrix. Second, we apply it to the components $\hat{F}_1$, $\hat{F}_2$, $\hat{F}_3$, and $\hat{F}_C$ individually and then form a block diagonal matrix out of the non-linearly shrunken components.

2.2.4 Robust time variation

As stock returns typically exhibit volatility clustering, we also consider a robustness analysis where we allow for time variation in the covariance matrices. We note that for monthly data volatility clustering is much less pronounced than for daily data. To track time-varying (co)variances, we consider a simple exponentially weighted moving average (EWMA) scheme. Let $\bar{f}$ represent the sample average of the (estimated) factor returns over the estimation window, then the EWMA recursion for $\hat{F}_t$ becomes

$$\hat{F}_{t+1} = \lambda \hat{F}_t + (1-\lambda)(\hat{f}_t - \bar{f})(\hat{f}_t - \bar{f})', \tag{9}$$

for $t = 1, \ldots, T$, where $\hat{f}_t - \bar{f}$ denotes the vector of estimated (de-meaned) factor returns (including currency returns), and where $0 < \lambda < 1$ denotes the decay factor. Similarly, for the variances of the idiosyncratic returns $\hat{\sigma}^2_{i,t}$, we follow a standard EWMA recursion as given by

$$\hat{\sigma}^2_{i,t+1} = \lambda \hat{\sigma}^2_{i,t} + (1 - \lambda)\hat{u}^2_t, \tag{10}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.[6] Throughout this paper we treat λ as a tuning parameter and set it to 0.97, which is the value typically used for a monthly frequency of observations.

Although an EWMA scheme can capture the time variation in the covariances, it is also subject to more estimation risk. As λ decreases, the end-of-sample covariance estimate puts more weight on a relatively small proportion of recent observations. The question of whether the EWMA yields better estimates is ultimately an empirical question. Therefore, we include a comparison of equally weighting and EWMA weighting in the empirical analysis later on.

Given the potential presence of outliers in stock returns, we also consider robustified EWMA schemes for the idiosyncratic variances $\hat{\sigma}^2_{i,t}$. The standard EWMA recursion in (10) is prone to outliers. These potentially have a substantial and long lasting impact on the estimated variance given the high value of λ. To robustify the methodology against such effects, we consider versions that are less sensitive to outliers. Our first robust time-varying approach uses the score-driven EWMA proposed in Lucas and Zhang (2016).[7] Here, the update is based on a scaled version of the score of the standardized Student's t distribution, i.e.,

$$\hat{\sigma}^2_{i,t+1} = \lambda \hat{\sigma}^2_{i,t} + (1 - \lambda)\frac{\tau + 1}{\tau - 2 + \hat{u}^2_t / \hat{\sigma}^2_{i,t}}\hat{u}^2_t. \tag{11}$$

[6] We initialize $\hat{F}_1$ and $\hat{\sigma}^2_{i,1}$ $\forall i$ using their sample covariances over the estimation window, i.e., $\hat{F}$ and $\hat{\sigma}^2_i$.

[7] This EWMA scheme is based on the score-driven volatility models proposed in Creal et al. (2011, 2013) and Harvey (2013).

We treat the degrees of freedom parameter τ as a tuning parameter and fix it at $\tau = 5$, which is a typical empirical value for stock returns; see Lucas and Zhang (2016)). Lower values of τ reduce the impact of outliers more strongly. Note that for $\tau \to \infty$ the EWMA scheme in (11) converges to the standard EWMA in (10).

Our final EWMA scheme is the robust EWMA recursion of Guermat and Harris (2002)

$$\hat{\sigma}_{i,t+1} = \lambda \hat{\sigma}_{i,t} + (1-\lambda)\sqrt{2}\,|\hat{u}_t|\,, \tag{12}$$

which is based on the robustified GARCH models of Taylor (1986) and Schwert (1990). It considers the absolute values of the idiosyncratic returns, rather than the squares, which makes it more robust to outliers. Unlike the score-driven EWMA specification in (11), however, large values of $\hat{u}_t$ can still have an unbounded impact on $\hat{\sigma}^2_{i,t+1}$.

3 Data and methodology

3.1 Data

We observe monthly stock returns from May 1999 until August 2017, covering 29 countries and 3895 different stocks in total. All stocks have been included in the MSCI World Index at least once during the sample period. The factor loadings X_t are taken from the Barra GEM2 model, which is one of the standard industry FFMs. See Menchero et al. (2008) for a detailed description regarding the construction of these factors. For each stock we have the exposure with respect to industry, country, and style factors as measured at the start of the month. We also observe the market capitalization for each stock and its inclusion status in the MSCI World Index (at the beginning of each month).

Given that our FFM also includes currency risk factors, we take the local returns for each stock. The exchange rate returns f_t^C are taken against the Euro for the Australian Dollar, the Canadian Dollar, the Swiss Franc, the

Danish Krone, the British Pound, the Hong Kong Dollar, the Japanese Yen, the Malaysian Ringgit, the Norwegian Krone, the New Zealand Dollar, the Singapore Dollar, the Swedish Krona, and the US Dollar. Finally, our proxy for the risk-free interest rate is the one-month Euro LIBOR rate.[8]

3.2 Methodology

To evaluate the effect of shrinkage and robust covariance matrix dynamics, we use a standard rolling estimation window of $T = 60$ months to estimate the parameters, apply shrinkage, and forecast the covariance matrix out-of-sample for month $T + 1$. We do so for each of the competing methods. For each method, we subsequently determine the global minimum volatility portfolio weights ω_{MV} implied by the forecast of the covariance matrix. The weights ω_{MV} follow from

$$\omega_{MV} = \arg\min_{\omega} \ \omega'\hat{\Sigma}\omega, \quad \text{subject to} \quad \sum_{i=1}^{N} \omega_i = 1, \tag{13}$$

which yields the analytical solution

$$\omega_{MV} = \hat{\Sigma}^{-1}\iota \ / \ \iota'\hat{\Sigma}^{-1}\iota, \tag{14}$$

where ι denotes a vector of ones. Our minimum volatility portfolios allow for both long and short positions. When imposing further short-sale and concentration constraints the analytical solution for ω_{MV} is not available anymore and numerical methods are required to obtain ω_{MV}. Given our vast dimension of $\hat{\Sigma}$ the resulting computational burden is very high. Therefore, we stick to the analytically tractable setting that allows for short positions. These additional constraints would typically only narrow the gap between the different methods.

The first (rolling) estimation window covers the period May 1999 until

[8] The exchange rates, LIBOR rates, market capitalizations, and the stock returns are obtained from the IDC database.

April 2004. We then compute the return of the minimum volatility portfolio over May 2004 for the competing methods. Next, we shift the estimation window one month forward and repeat the whole procedure until we reach August 2017, which is the last month in our evaluation sample. The total evaluation period covers 160 out-of-sample months for which we have the returns of the minimum volatility portfolios. We evaluate the competing methods by the realized volatility $\hat{\sigma}_{MV}$ of these returns. A lower value of $\hat{\sigma}_{MV}$ indicates a better performance of the underlying method. Next to $\hat{\sigma}_{MV}$, we also consider the mean return $\hat{\mu}_{MV}$ and the Sharpe ratio

$$\hat{S}_{MV} = \hat{\mu}_{MV} \;/\; \hat{\sigma}_{MV}. \tag{15}$$

Although $\hat{\mu}_{MV}$ and $\hat{S}_{MV}$ are unrelated to the minimum volatility objective function underlying ω_{MV}, these measures provide a useful complementary perspective on the out-of-sample performance of the portfolio returns.

We also evaluate the properties of the minimum volatility portfolio weights ω_{MV} in more detail. First, we measure the effective number of stocks (ENS) in the minimum volatility portfolio, defined as

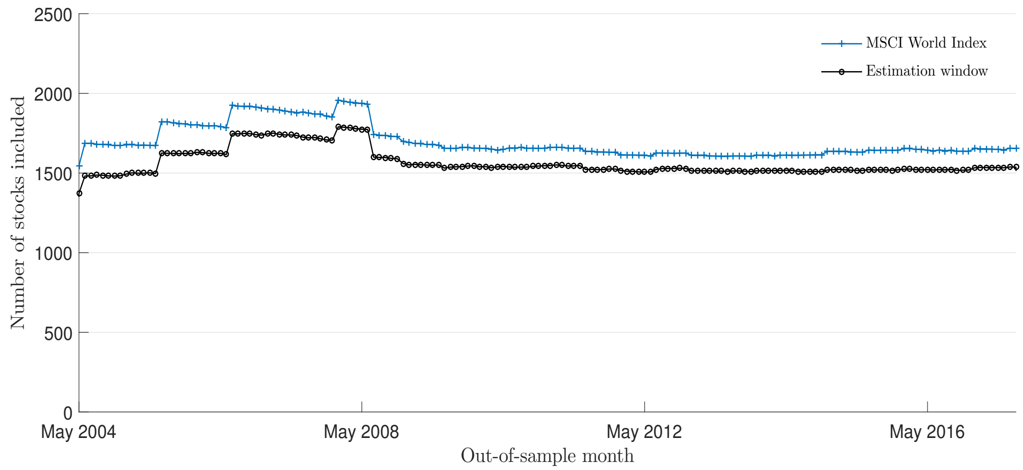
$$ENS = \left(\sum_{i=1}^{N} |\hat{\omega}_i| \right)^2 \Bigg/ \sum_{i=1}^{N} \hat{\omega}_i^2. \tag{16}$$

This is a standardized inverse Herfindahl-Hirschman index, where we standardize the portfolio weights $\hat{\omega}_i$ by $\sum_{i=1}^{N} |\hat{\omega}_i|$ to account for possible short positions in a selection of the stocks. In general, a higher ENS value implies a more diversified portfolio. As an alternate measure of riskiness, we also consider the total short position (TSP) for each portfolio, measured as

$$TSP = -\sum_{i=1}^{N} \hat{\omega}_i \cdot I\{\hat{\omega}_i < 0\}, \tag{17}$$

where $I\{\hat{\omega}_i < 0\}$ is 1 if $\hat{\omega}_i < 0$, and zero else. For both measures we report the averages over the 160 months in the evaluation period.

Figure 1: Stock inclusion



Notes: The figure displays for each month in the evaluation sample the number of stocks included in the MSCI World Index together with the number of stocks included in our estimation window.

To mimic the set of stocks in the MSCI World Index as closely as possible, for each estimation window we include in our sample those stocks that are part of the MSCI World Index at the end of the last month and for which we have 60 past return observations and one future return observation. The latter results in a small look-ahead-bias. We do not expect this to affect our conclusions in any material way. Furthermore, some stocks drop out since they belong to industries or countries for which the factor returns are not available. For a specific industry or country factor to be included, we require at least five stocks in that category. Otherwise, there is not enough information to obtain a sufficiently accurate estimate of the corresponding factor return and the corresponding stocks are left out of the rolling estimation sample.

Figure 1 compares the number of stocks in the MSCI World Index with the number of stocks included in our estimation windows for the 160 out-of-sample months. There is only a relatively small gap between the set of stocks included in our estimation samples and those included in the MSCI World Index, ranging from a reduction of 13% in the number of firms at the start of the sample, to a low 6% towards the end of the sample. In any case, there is no reason to expect sample selection issues that would put the

Table 1: Included factors over the estimation windows

	average	minimum	maximum
# country factors	21.73	20	23
# industry factors	33.86	32	34
# style factors	9.00	9	9
# exchange rates	11.80	11	12
# total	76.39	74	78

Notes: The table gives a summary of the number of included factors (per category) over the 160 estimation windows.

competing covariance estimation methods and shrinkage approaches on an unequal footing in the empirical analysis in Section 4.

Table 1 summarizes the number of included factors over the 160 estimation windows. The factors are split out over the four different categories, namely country factors, industry factors, style factors, and exchange rates. Factors may drop out during a particular period if there are not enough stocks in a particular country or industry to obtain a sufficiently reliable estimate of the fundamental factor return. As illustrated by Table 1, there is only a minor variation over time in the set of included factors, again underlining that sample selection issues do not play a major role in our analysis.

Our sample of stocks is based on the MSCI World Index, which consists of stocks with a large market capitalization. For this reason we do not find substantial outliers in the data, unlike a situation where micro-caps are included in the sample. We find some large returns, positive as well as negative. The minimum return over the entire sample was around -87%, while the maximum was around 264%. In most months, however, the returns were between -50% and 100%, suggesting there will be no problems with outliers or influential observations in our analysis.

3.3 Competing estimation and shrinkage methods

Our base case is the FFM with the factor return covariance matrix F and idiosyncratic covariance matrix U estimated by their standard sample covariance matrices. To account for the estimation uncertainty in these covariance matrices that make up the covariance matrix of asset returns, we consider six alternatives where we shrink either the sample covariance matrix $\hat{F}$ of the fundamental factor returns or the sample idiosyncratic covariance matrix $\hat{U}$. The first two methods use linear shrinkage of $\hat{F}$. First, we consider the shrinkage target $\hat{F}^0$ as in (5) with equicorrelation within each group of factor returns. The second shrinkage method goes one step further and uses a target with zero correlations. Methods three and four use the non-linear shrinkage of Ledoit and Wolf (2017a) rather than linear shrinkage. Method three uses non-linear shrinkage on the full $\hat{F}$, whereas method four considers non-linear shrinkage of the four components in (5) individually. Method five is set up to check the relevance of the estimated correlations in the entire composite covariance matrix $\hat{\Sigma}$ of the FFM. It dogmatically shrinks $\hat{\Sigma}$ to a diagonal target $\hat{\Sigma}^0$ that has the same diagonal as $\hat{\Sigma}$ by setting the off-diagonal elements to zero. Finally, method six does not shrink the sample factor covariance matrix, but instead considers linear shrinkage on the idiosyncratic variances to a target $\hat{U}^0$ that has a common idiosyncratic variance equal to the average idiosyncratic variance in the original $\hat{U}$.

Rather than shrinking the components $\hat{F}$ and $\hat{U}$ of the covariance matrix $\hat{\Sigma}$, we can also shrink the sample covariance matrix directly without imposing the FFM structure. We do so in the next four methods. The first and second of these shrink the sample covariance matrix of returns linearly towards an equicorrelation and a zero-correlation target, respectively. Note that the sample covariance matrix itself cannot be used for determining the minimum variance portfolio, as the number of observations in the estimation sample ($T = 60$) is much smaller than the number of assets ($N > 1,500$). This results in a non-unique solution to the optimization problem. Such problems are avoided by the linear shrinkage techniques. We compute the shrinkage constant for these two methods based on Ledoit and Wolf (2004a). The third approach within this group uses the non-linear shrinkage method of Ledoit and Wolf (2017a) on the sample covariance matrix. This, actually, would be the most appealing as it avoids imposing the fundamental factor structure, while still eliminating the estimation bias of the full sample covariance matrix. Good results with this approach were reported in Engle et al. (2017). Our fourth method is a much more blunt benchmark, where we put all off-diagonal elements of the sample covariance matrix to zero. It results in an equal risk weighted portfolio when determining the minimum variance portfolio.

The final two benchmarks that we include in our analysis do not build on the estimated covariance matrix in any way, but use pre-determined portfolio weights, namely market cap based weights, and equal weights. Particularly the equal weigthing provides a naive, but typically rather strong benchmark in the presence of estimation error; see DeMiguel et al. (2007).

4 Results

In this section we present the empirical results. We first discuss the out-of-sample minimum variance portfolio performance using the different shrinkage methods for the FFM. We then discuss the choice of the shrinkage constant. Finally, we investigate the effect of allowing for time variation in the FFM's covariance matrix.

4.1 Shrinkage

Table 2 presents the results. We first focus on the FFM results. The standard FFM without any shrinkage already provides a strong benchmark in that its minimum variance portfolio returns have the second lowest out-of-sample standard deviation. Its average return is correspondingly lower, but its Sharpe ratio is still in the top three. All shrinkage methods for the covariance matrix $\hat{F}$ appear ineffective in lowering the out-of-sample minimum variance portfolio standard deviation. This holds both for the linear and non-linear shrinkage techniques, as well as for shrinkage targets with equicorrelation, zero correlation, or with a block structure. The bias effect caused

Table 2: Minimum volatility portfolio returns

Details	Std. dev.	Mean	Sharpe	ENS	TSP
FFM- standard	8.00	8.83	1.10	692.88	1.57
FFM- shrinkage: equicorrelation	8.81	9.56	1.08	649.08	0.99
FFM- shrinkage: zerocorrelation	8.57	9.78	1.14	671.52	0.97
FFM- nonlin shrinkage: full $\hat{F}$	9.41	9.70	1.03	650.98	1.19
FFM- nonlin shrinkage: parts $\hat{F}$	8.93	9.04	1.01	711.14	1.13
FFM- diagonal $\hat{\Sigma}$	12.56	8.52	0.68	1160.83	0.00
FFM- shrinkage: idiosyncratic	7.67	8.79	1.15	841.88	1.43
Sample- shrinkage: equicorrelation	8.21	9.18	1.12	536.20	1.05
Sample- shrinkage: zerocorrelation	7.38	7.36	1.00	562.73	0.62
Sample- nonlin shrinkage: full $\hat{\Sigma}$	8.20	7.14	0.87	1005.33	0.56
Sample- diagonal $\hat{\Sigma}$	12.01	8.54	0.71	945.67	0.00
Benchmark- marketcap weighted	13.08	7.02	0.54	336.60	0.00
Benchmark- equally weighted	14.71	8.24	0.56	1564.96	0.00

Notes: The table displays the out-of-sample evaluations of the returns of the minimum volatility portfolios obtained from the competing estimation methods. The evaluation period runs from May 2004 until August 2017 and consists of 160 months. Furthermore, Sharpe represents the estimated Sharpe ratio, ENS the effective number of stocks, and TSP the total short position, see Section 3.2 for more details on these evaluation measures.

by shrinkage therefore seems to off-set any potential gains in estimation error reduction when considering the estimation of $\hat{F}$.

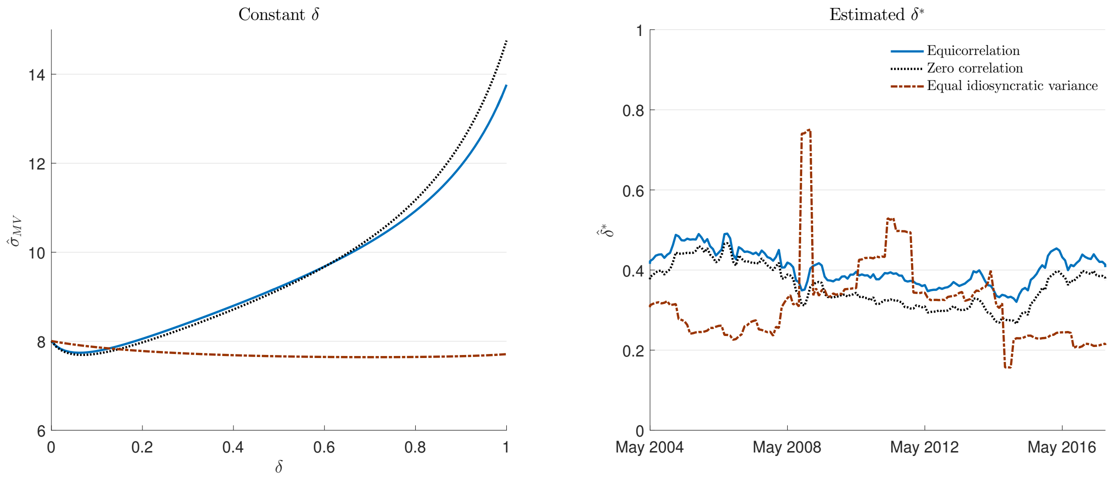
The only effective shrinkage method for the FFM appears to be the shrinkage of the idiosyncratic (diagonal) covariance matrix $\hat{U}$. Shrinkage of this matrix to a matrix with equal idiosyncratic variances produces the lowest ex-post standard deviation of 7.67 per cent. We note that given our relatively short out-of-sample period, the standard errors for the standard deviations, means, and Sharpe ratios in Table 2 are too high to make statements about statistical significance of the differences between models. Still, it is interesting to see that the estimation error in the idiosyncratic variances appears to be substantial, and that shrinkage on these idiosyncratic variances improves both the standard deviation and the Sharpe ratio. Interestingly, shrinking $\hat{U}$ results in a higher effective number of stocks (ENS) and a somewhat smaller total short position (TSP) than the standard fundamental factor model. The FFM with shrunken $\hat{U}$ thus also appears to yield a slightly more balanced and less leveraged minimum volatility portfolio.

The importance of the correlations between stock returns implied by the FFM structure is also evident from the return performance of the FFM with diagonal $\hat{\Sigma}$ shrinkage target. Here, we use the variances of the FFM, but put all correlations in the shrinkage target to zero. This method is the worst performer in the set of models considered. It even performs worse than the diagonal counterpart of the sample covariance matrix (sample - diagonal $\hat{\Sigma}$). It thus appears important to include the estimated correlations in the portfolio optimization, despite the estimation uncertainty caused by small samples (60 months) and a large number of stocks ($N > 1,500$).

If we abandon the FFM covariance structure and shrink the complete sample covariance matrix $\hat{\Sigma}$ directly, we see a number of interesting features. First, the non-linear shrinkage techniques of Ledoit and Wolf (2015, 2017a) when applied to $\hat{\Sigma}$ appear less effective than a standard FFM, both in terms of the ex-post standard deviation and Sharpe ratio. This holds despite the number of effective stocks (ENS) in this portfolio being considerably higher than in the standard FFM, and the total short position (TSP) being lower. This performance is somewhat disappointing given the promising character of non-linear shrinkage in vast dimensions. The linear shrinkage techniques applied to the sample covariance matrix result in a comparable performance of the minimum variance portfolio to the FFM based portfolios. Linear shrinkage of the sample covariance matrix to a zero correlation shrinkage target even appears to be the best performer in terms of ex-post minimum standard deviation amongst all estimators considered. The method performs somewhat less, however, in terms of Sharpe ratio and effective number of stocks.

To conclude the discussion of Table 2, we note that both the market cap weighted portfolio and the equally weighted portfolio have a substantially higher out-of-sample standard deviation than any of the shrinkage estimation techniques. Also the Sharpe ratios are correspondingly lower. It therefore pays off in portfolio optimization to measure the variance of the stock returns and account for their dependence structure in vast dimensions.

Figure 2: Shrinkage intensities



Notes: The left panel shows for a grid of fixed δ values the resulting volatility of the minimum volatility portfolio. The right panel displays the estimated δ values based on the estimation window.

4.2 Linear shrinkage intensity

The linear shrinkage coefficient δ was determined for each estmation window in the analysis in Section 4.1 using the methods of Ledoit and Wolf (2003, 2004b). The precise values of δ of course affects the relative performance of the different methods. To shed some light on this sensitivity, we first compute the out-of-sample performance of the minimum variance portfolio for a grid of δ, whereby we fix δ to be constant over the estimation windows. We then evaluate the ex-post standard deviation $\hat{\sigma}_{MV}$ as a function of δ, and plot the result in the left-hand panel of Figure 2. We use two (linear) shrinkage targets, namely a factor covariance matrix with an equicorrelation and a zero correlation assumption. We also include the shrinkage estimator with equal idiosyncratic variances ($\hat{U} = \bar{\sigma}^2 \mathrm{I}$) into our comparison.

The left-hand panel in Figure 2 reveals that improvements over the standard FFM are possible for low shrinkage intensities towards an equi- or zero correlation target. The gains are modest and of the order of 3% or less. Moreover, the gains are quickly lost and turn into losses for shrinkage intensities δ above 0.2. Linear shrinkage towards a $\hat{U}$ with equal variances, however, is much more robust. Though the gains are only slighly larger than those of the equicorrelation and zero correlation shrinkage of $\hat{F}$, the gains appear much less sensitive to the precise value of δ. Almost all gains are even still realized for full shrinkage, i.e., an extreme value of $\delta = 1$.

Given the sensitivity of the results to the coefficient δ, in particular for shrinkage of $\hat{F}$, it is important to consider the plug-in values of the shrinkage coefficient δ in the empirical results. The right-hand panel of Figure 2 plots the estimated values of δ resulting from the estimation windows as a function of time. We see that for both shrinkage targets of $\hat{F}$, the parameter δ is estimated in the range $0.3 - 0.5$ using the approach of Ledoit and Wolf (2003). This explains why the linear shrinkage estimators of $\hat{F}$ perform poorly in Table 2. The plug-in estimator of δ looses (or actually reverses) the possible shrinkage gains.[9]

Finally, the right-hand plot in Figure 2 also shows the plug-in value of δ for the shrinkage target of $\hat{U}$ with equal idiosyncratic variances. This value of δ varies much more strongly over time. There is a clear peak directly after the financial crisis, resulting in more shrinkage being imposed. As the gains in terms of the minimum volatility portfolio performance σ_{MV}, however, is much less sensitive to the precise value of δ, the strong time series variation in δ in this case does not affect the fact that this shrinkage target realizes the gains compared to the standard FFM as evidenced in Table 2.

4.3 Time variation

We now turn to the possibility of time variation in the factor covariance matrix F_t and the idiosyncratic covariance matrix U_t. As our benchmark, we take the standard FFM with time-invariant F and U. We consider three settings. The settings differ in the way the idiosyncratic covariance matrix U_t is made time-varying. We use the standard EWMA and the two robust schemes of Section 2, respectively. All three settings use a standard EWMA

[9] Note that this estimator is specified to minimize the Frobenius norm in (7). This criterion does not necessarily imply optimality with respect to the implied minimum volatility portfolio weights.

Table 3: Minimum volatility portfolio, time variation in $\hat{F}$ and $\hat{U}$

time variation $\hat{F}_t$	time variation $\hat{U}_t$	Std. dev.	Mean	Sharpe	ENS	TSP
none	none	8.00	8.83	1.10	692.88	1.57
EWMA	EWMA	8.04	8.88	1.10	694.20	1.58
EWMA	student-t score	8.02	8.96	1.12	706.20	1.53
EWMA	absolute returns	7.98	8.92	1.12	702.14	1.51

Notes: The table displays the out-of-sample evaluations of the returns of the minimum volatility portfolios obtained from the competing EWMA schemes.

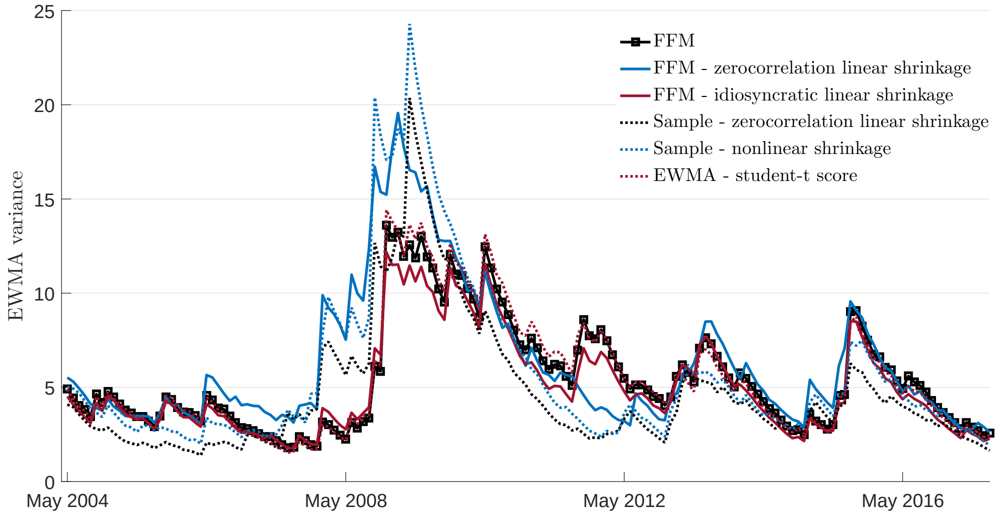
scheme to estimate F_t. Table 3 presents the results.

We conclude that allowing for time variation in our setting does not have a major impact. There are only mild changes in the minimum volatility return performance for the three methods considered compared to the standard FFM setting. Also the portfolio weight measures ENS and TSP are hardly effected. The exponential decay of the weights for the returns over the limited length of the estimation period of 60 months apparently does not yield a more accurate covariance matrix estimate compared to the setting of using equal weigths over the estimation sample, as is done for the standard FFM. Note that this result is not at odds with the typical finding of volatility clustering in stock return data. First, the return data are observed at a monthly rather than a daily frequency, which already mitigates volatility clustering effects. Second, the FFM's structure already implies a time-varying covariance matrix through the time variation in the factor loadings X_t. Finally, the short rolling estimation sample of 60 months also mimics part of the volatility clustering effects present at a lower frequency and therefore further mitigates the potential gains of more complex methods such as EWMA, robust filters, or GARCH type models. The additional gains of time-varying parameter models, therefore, appear less useful in the current setting.

4.4 Time series pattern of the minimum volatility

The preceding results in Table 2 provided the realized volatility $\hat{\sigma}_{MV}$ of the minimum volatility portfolio returns over the entire out-of-sample period. In

Figure 3: Minimum variance portfolio performance over time



Notes: The figure displays the locally estimated variance of the minimum volatility returns. These are obtained through an EWMA scheme with decay parameter $\lambda = 0.9$.

Figure 3, we show the time series pattern of $\hat{\sigma}_{MV}$ using a simple EWMA smoothing scheme using a smoothing parameter $\lambda = 0.9$, which corresponds to a half-life of 6 or 7 months.

The figure reveals a number of interesting features. First, the time series pattern in Figure 3 shows that the averaged results in Table 2 are not limited to an isolated period, but actually persist over longer episodes. For instance, the standard FFM has a stable low-volatility performance, only improved by the shrinkage method with equal idiosyncratic variances. The difference between the latter two is generally very small, except during the great financial crisis and early 2012. Second, the linear shrinkage methods often perform at par with the standard FFM, except in the run-up of the great financial crisis. There, the omission in the shrinkage of the correlation effects or their heterogeneity results in much higher ex-post variances. This makes sense, as correlations typically increase over crisis periods. We also find that introducing (robust) time variation only has a minor impact on the minimum volatility portfolio variance, irrespective of the period. The EWMA results are hardly distinguishable from the standard FFM. Finally, the non-linear shrinkage of the standard sample covariance matrix does not appear to work well in the run-up towards the crisis. The ex-post minimum volatility portfolio variances based on the non-linearly shrunken sample covariance matrix can be more than twice the size of its counterpart based on the standard FFM. Only in 2005-2006 and in 2012, the non-linear shrinkage method results in a better performance than the FFM. As these appear to be rather low variance episodes, these gains do not seem to off-set the increased variance effects during the turbulent crisis period.

5 Conclusion

In this paper, we focused on estimation risk and statistical shrinkage methods in vast-dimensional fundamental factor models (FFMs). The FFM provides a convenient framework to estimate a covariance matrix of equity returns, even for a large number of stocks, and is the standard in the asset management industry. In order to address the issues caused by having to estimate a vast number of parameters based on a limited estimation sample, we considered the potential gains from statistical shrinkage techniques in FFM covariance matrices in an out-of-sample context. As our evaluation criterion we used the ex-post volatility of the minimum volatility portfolio returns.

Using a large sample of stocks that mimics the MSCI World Index, we found that the standard FFM already shows a good performance. Linear and non-linear shrinkage techniques applied to the fundamental factor returns covariance matrix resulted in a worse performance during almost all periods. When zooming in on the determination of the shrinkage coefficient, we found that standard plug-in values for this coefficient worsened rather than improved the situation. This analysis further indicated that the potential empirical gains from shrinking the factor returns covariance matrix are quite limited. Better and more robust results were obtained with shrinkage targets based on the idiosyncratic covariance matrix. Also there, however, the gains were limited and not statistically significant. Allowing for time variation in the underlying factor return covariance matrix or the idiosyncratic covariance matrix did not alter our results.

A final interesting finding emerged from the application of the recent non-linear shrinkage techniques of Ledoit and Wolf (2015, 2017a) applied to the entire sample covariance matrix. Also these techniques do not seem to improve upon the FFM performance. By contrast, the ex-post variance of the minimum volatility portfolio turned out the be sometimes more than twice as high during the turbulent period around the great financial crisis. The lower variance of non-linear shrinkage methods during calmer periods did not off-set this.

In conclusion, our results suggest that estimation risk in stock return correlations in vast-dimensional spaces can be adequately addressed using fundamental factor models. Linear shrinkage techniques with standard plug-in tuning parameters typically deteriorate rather than improve the situation, as do non-linear shrinkage methods. Our analysis further indicated that the potential gains of shrinkage methods compared to the standard FFM are quite modest, leaving the FFM as a typical benchmark model in vast-dimensions.

References

Brito, D., M. C. Medeiros, and R. Ribeiro (2018). Forecasting large realized covariance matrices: The benefits of factor models and shrinkage.

Brown, S. J. (1989). The number of factors in security returns. *the Journal of Finance 44*(5), 1247–1262.

Connor, G. and R. A. Korajczyk (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of financial economics 15*(3), 373–394.

Creal, D., S. J. Koopman, and A. Lucas (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics 29*(4), 552–563.

Creal, D., S. J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics 28*(5), 777–795.

DeMiguel, V., L. Garlappi, and R. Uppal (2007). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies 22*(5), 1915–1953.

Engle, R. F., O. Ledoit, and M. Wolf (2017). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 1–13.

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy 81*(3), 607–636.

Guermat, C. and R. D. Harris (2002). Robust conditional variance estimation and value-at-risk. *Journal of Risk 4*, 25–42.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*, Volume 52. Cambridge University Press.

Lam, C. et al. (2016). Nonparametric eigenvalue-regularized precision or covariance matrix estimator. *The Annals of Statistics 44*(3), 928–953.

Ledoit, O. and M. Wolf (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 M. Wolf (2004a). Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management 30*(4), 110–119.

Ledoit, O. and M. Wolf (2004b). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis 88*(2), 365–411.

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

Ledoit, O. and M. Wolf (2017a). Direct nonlinear shrinkage estimation of large-dimensional covariance matrices.

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

Ledoit, O. and M. Wolf (2018). Optimal estimation of a large-dimensional covariance matrix under stein's loss. *Bernoulli 24*(4B), 3791–3832.

Lucas, A. and X. Zhang (2016). Score-driven exponentially weighted moving averages and value-at-risk forecasting. *International Journal of Forecasting 32*(2), 293–302.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance 7*(1), 77–91.

Menchero, J., A. Morozov, and P. Shepard (2008). The barra global equity model (gem2). *MSCI Barra Research Notes*, 53.

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

Schwert, G. W. (1990). Stock volatility and the crash of'87. *The Review of Financial Studies 3*(1), 77–102.

Taylor, S. J. (1986). *Modelling financial time series.* world scientific.

Appendix A Shrinkage intensity

The linear shrinkage methods in the FFM require us to set a shrinkage coefficient. Here, we provide the formulas for the estimator of the optimal shrinkage coefficient. We first consider the covariance matrix $\hat{F}$ of fundamental factor returns. Then, we consider the matrix $\hat{U}$ holding idiosyncratic variances.

Factor covariance matrix

Recall that the estimated fundamental factor and currency returns are represented by $\hat{f}_{i,t}$, for $i = 1, \ldots, K^*$ and $t = 1, \ldots, T$. Furthermore, let $\hat{F}$ denote the sample covariance matrix, and $\hat{F}^0$ the shrinkage target. The true covariance matrix of the factor returns is represented by F. In general, we want to use the optimal shrinkage value δ^* defined by

$$\delta^* = \text{argmin}_{\delta \in [0,1]} \mathbb{E}\left[||(1-\delta)\hat{F} + \delta \hat{F}^0 - F||^2 \right]. \tag{A1}$$

Ledoit and Wolf (2003) show that the optimal δ^* behaves like a constant over the sample size T as $T \to \infty$ (for fixed K^*, up to higher-order terms of T), i.e. $\delta^* \approx \kappa/T$ for large T. This constant κ can be expressed as

$$\kappa = \frac{\pi - \phi}{\gamma}, \tag{A2}$$

a description of π, ϕ, and γ can be found in Ledoit and Wolf (2003). For our equicorrelation shrinkage target, the formulas are derived in Ledoit and Wolf (2004a). We repeat these here. We use the subscript i, j to indicate the element in row i and column j of a matrix. A consistent estimator for π is given by

$$\hat{\pi} = \sum_{i=1}^{K^*} \sum_{j=1}^{K^*} \hat{\pi}_{i,j}, \quad \text{with} \quad \hat{\pi}_{i,j} = \frac{1}{T} \sum_{t=1}^{T} \left((\hat{f}_{i,t} - \overline{f}_{i,\cdot})(\hat{f}_{j,t} - \overline{f}_{j,\cdot}) - \hat{F}^0_{i,j} \right)^2, \tag{A3}$$

where $\overline{f}_{i,\cdot}$ represents the average return for factor i. Similarly, γ can be estimated consistently using

$$\hat{\gamma} = \sum_{i=1}^{K^*} \sum_{j=1}^{K^*} \left(\hat{F}^0_{i,j} - \hat{F}_{i,j} \right)^2 . \tag{A4}$$

The estimators for π and γ do not depend on the specific form of the shrinkage target $\hat{F}^0$. Instead, the estimator for ϕ is specific to the shrinkage target. For the case of an equicorrelation shrinkage target,Ledoit and Wolf (2004a) show that ϕ can be consistently estimated using

$$\hat{\phi} = \sum_{i=1}^{K^*} \hat{\pi}_{i,i} + \sum_{i=1}^{K^*} \sum_{j \neq i} \frac{\hat{\rho}}{2} \left(\sqrt{\frac{\hat{F}_{j,j}}{\hat{F}_{i,i}}} \hat{\nu}_{ii,ij} + \sqrt{\frac{\hat{F}_{i,i}}{\hat{F}_{j,j}}} \hat{\nu}_{jj,ij} \right), \tag{A5}$$

where $\hat{\rho}$ represents the average of the estimated correlations in the sample covariance matrix $\hat{F}$, and

$$\hat{\nu}_{ii,ij} = \frac{1}{T} \sum_{t=1}^{T} \left((\hat{f}_{i,t} - \overline{f}_{i,\cdot})^2 - \hat{F}_{i,i} \right) \left((\hat{f}_{i,t} - \overline{f}_{i,\cdot})(\hat{f}_{j,t} - \overline{f}_{j,\cdot}) - \hat{F}_{i,j} \right), \tag{A6}$$

$$\hat{\nu}_{jj,ij} = \frac{1}{T} \sum_{t=1}^{T} \left((\hat{f}_{j,t} - \overline{f}_{j,\cdot})^2 - \hat{F}_{j,j} \right) \left((\hat{f}_{i,t} - \overline{f}_{i,\cdot})(\hat{f}_{j,t} - \overline{f}_{j,\cdot}) - \hat{F}_{i,j} \right). \tag{A7}$$

For the shrinkage target with zero correlation, we apply the formulas above and set $\hat{\rho} = 0$.

Given $\hat{\pi}$, $\hat{\phi}$, and $\hat{\gamma}$ we obtain $\hat{\kappa}$ from

$$\hat{\kappa} = \frac{\hat{\pi} - \hat{\phi}}{\hat{\gamma}}. \tag{A8}$$

The estimate of the optimal shrinkage intensity $\hat{\delta}^*$ is

$$\hat{\delta}^* = \max \left[0, \min \left[\frac{\hat{\kappa}}{T}, 1 \right] \right]. \tag{A9}$$

Idiosyncratic variances

Let the diagonal N-by-N matrix $\hat{U}$ represent the sample covariance matrix of the estimated idiosyncratic returns $\hat{u}_{i,t}$, for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. Furthermore, let $\bar{\sigma}^2$ denote the average over the N estimated idiosyncratic variances. Ledoit and Wolf (2004b) show that the optimal shrinkage parameter is given by

$$\delta^* = \frac{b^2}{d^2}. \tag{A10}$$

They show that a consistent estimator for d^2 is given by

$$\hat{d}^2 = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{U}_{i,i} - \bar{\sigma}^2 \right)^2, \tag{A11}$$

and for b^2 by

$$\hat{b}^2 = \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \left((\hat{u}_{i,t} - \overline{u}_{i,\cdot})^2 - \hat{U}_{i,i} \right)^2, \tag{A12}$$

where $\overline{u}_{i,\cdot}$ represents the average (estimated) idiosyncratic return for stock i. The estimated optimal shrinkage intensity $\hat{\delta}^*$ is then given by

$$\hat{\delta}^* = \frac{\min\left[\hat{d}^2,\ \hat{b}^2\right]}{\hat{d}^2}. \tag{A13}$$