Optimizing large portfolios using factors and sparse hedging

Rasmus Lönn*

October, 2020

Abstract

Optimizing portfolios of many assets requires an estimate of a high-dimensional inverse covariance matrix. In this paper we promote an estimation approach that exploits a combination of financial factors and sparsity restrictions. Whereas prominent existing methods impose sparsity restrictions in the covariance matrix we argue that these constraints are better suited for portfolio allocation if imposed directly on the inverse covariance matrix. Such constraints are financially corresponding to regularized hedging portfolios, which can be estimated using the graphical lasso. In combination with financial factors this approach also circumvents the explicit high-dimensional matrix inversion. Empirically, we find that this method tends to reduce portfolio volatility more than both competing linear and non-linear shrinkage methods.

Keywords: High-dimensionality; Asset allocation; Graphical model selection

JEL codes: C55 (Large Data Sets), G11 (Portfolio Choice)

* Econometric Institute, Erasmus School of Economics (ESE), Erasmus University, P.O. Box 1738, 3000DR Rotterdam, The Netherlands. Email: lonn@ese.eur.nl

Thanks to seminar participants at NOVA School of Business and Economics, University of Konstanz, Maastricht University, Radboud University Nijmegen, Erasmus University Rotterdam and conference participants at the Workshop on Financial Econometrics 2018 (Örebro) and Twelfth Annual SoFiE Conference 2019 (Shanghai), for helpful comments and suggestions.

1 Introduction

Uncertainty in the estimates of the first and second moments of returns reduces the out-of-sample performance of the fundamental Markowitz (1952) portfolios. These issues are critical when portfolios grow to include more assets. However, with the many improvements in the estimation of high-dimensional covariance matrices and their inverses this is becoming a viable challenge. In this paper we further contribute to this literature by exploring the use of sparsity restrictions when factor models fail to capture all co-variation in returns. We show that the manner in which sparsity constraints are imposed matter greatly for portfolio optimization. In particular, whereas prominent estimators impose sparsity in the pairwise relations of returns, we find that simple financial models favour sparsity in the *partial* linear relations. Our empirical study concurs with this finding. The method we promote uses the graphical lasso estimator conditional on a factor model of observed financial factors. We consider both the case of *a priori* imposed factors as well as the case where factors have to be selected from a large number of candidates. This introduces an interesting second model selection challenge, which can also be approached using sparsity restrictions. We depart from previous works in the financial literature by emphasizing the benefits of conditionally sparse system of partial relations and their appeal under common financial return structures.

It is well known that mean-variance optimal portfolios despite being theoretically appealing often perform poorly out-of-sample (Jobson and Korkie, 1980; Michaud, 1989) and DeMiguel et al. (2007) show how many estimators fail to outperform an equally weighted allocation. Restricting attention to minimum-variance allocation often alleviates the issues by reducing the estimation to only include the covariance matrix of returns. However, as the number of assets N grows large this no longer circumvents the main empirical challenge (Kan and Zhou, 2007). In these large-N allocations the number of parameters to estimate in the covariance matrix becomes vast, $N(N+1)/2$. In addition, the allocation rule also requires that the large-N covariance matrix be inverted.

There are several methods that approach these issues to date, prominent among them are linear and non-linear shrinkage estimators of the covariance matrix, Ledoit and Wolf (2003, 2004) and Ledoit and Wolf (2012, 2015, 2017) respectively. Shrinking the sample covariance matrix towards a well-conditioned target matrix gives an estimate, which produces a valid

estimate of the inverse covariance matrix. The optimal degree of shrinkage is estimated such that the estimates retain appropriate properties even in large-N applications. Ledoit and Wolf (2017) show how particularly non-linear methods are well-suited for portfolio optimization and how they may be specifically optimized for this task.

Appealing targets for shrinkage estimators can be drawn from the big literature on expected returns. These representations may be economic, or financial, models such as the CAPM (Sharpe, 1964), Fama-French models (Fama and French, 1993, 2015), or macro factors as in Chen et al. (1986). Alternatively, factors can stem from a statistical perspective such as those of Ross (1976) and Chamberlain and Rothschild (1983). Building on these factor representations, Fan et al. (2008) explore the asymptotic properties of the implied covariance matrices in high-dimensional settings. Further, Fan et al. (2013) show how a combination of non-observed factors and sparse residual relations can inform portfolio optimization. Decomposing the covariance matrix based on its k leading eigenvalues, the structure implied by the k eigenvalues is unchanged while a soft thresholding rule is applied to elements of the residual covariance matrix. Thresholding imposes sparsity in the residual covariance structure by adjusting some elements to zero. In addition, the sparse residual covariances enable deviations from the strict factor structure imposed by the leading k eigenvectors. In optimizing a portfolio this estimate is inverted.

An important issue for these estimators is the economic loss that follows when factor models may be misspecified. In this paper we show that when the residual covariances matrix admits an approximately equicorrelational pattern, we can benefit from imposing sparsity restrictions in the inverse covariance matrix (precision matrix), rather than the covariance matrix of the residual returns. This builds upon approaches such as Goto and Xu (2015), Cai et al. (2019) and Callot et al. (2019) that estimate the inverted matrix directly, thus circumventing the high-dimensional matrix inversion. In particular we focus on the penalized likelihoods of Yuan and Lin (2007), optimized using the graphical lasso (Friedman et al., 2008). The graphical lasso regularizes the estimate of the precision matrix by shrinking all entries and setting small elements equal to zero, thus using shrinkage to promote sparsity. It accomplishes this by regularizing the estimate of the off-diagonal element in the inverse under the absolute-value norm.[1]

[1]Cai et al. (2019) explores a related high-frequency setting, estimating the inverse directly using the

These methods impose that, conditional on a parsimonious model of factors, the precision matrix of returns is well represented by a sparse matrix. This form of conditional sparsity admits clear economic intuitions for portfolio management. The sparsity implies that only a small number of off-diagonal elements in the conditional precision matrix are non-zero. The implications of this assumption can be clarified through a regression perspective. For every asset return, y_i, the row/column i of the precision matrix is proportional to the negative regression parameters of $y_i = \sum_{j \neq i} y_j \gamma_j + \varepsilon_i$. Thus, the sparsity assumption implies that these regressions, for all i, are parsimonious models. Financially, this restriction can be understood as a statement about optimal hedging portfolios since the scaled negative regression coefficients correspond to optimal hedging weights (Stevens, 1998). *E.g.* in a portfolio of N assets, a long position in asset i is optimally hedged by a short position proportional to γ_j in asset j. To promote this sparsity in these hedging portfolios we add a few factors to the hedge regression. Importantly, the sparsity assumption is imposed on the partial correlations between assets not the pair-wise correlations. We find that this distinction is important once the parsimonious factor model may be misspecified.

When factors models are well specified such that the residual covariance matrix is diagonal covariance sparsity and precision sparsity restrictions are equivalent. We rather contrast these restrictions when residuals admit some omitted structure, in this paper we look at approximately equicorrelational residuals. We find that under such structures the relative volatility of the minimum variance allocation under covariance sparsity restrictions exceeds that of the precision sparse models. This finding leads us to an approach that targets the inverse directly while also exploiting a large number of factors in a manner that remains empirically viable when N is very large. The graphical lasso is especially well suited for this task since the regularization can be optimized to ensure estimates that are symmetric and positive semi-definite. Furthermore, it is straightforward to use cross-validation to calibrate the regularization in order to specifically minimize portfolio volatility. Financially, the graphical lasso performs hedging portfolio selection where assets are selected to complement the factor structure. This role is analogous to how the POET and some Shrinkage estimators enable deviations from the strict factor representation.

We evaluate the portfolio performance using daily returns of the S&P500 companies be-

CLIME estimator of Cai et al. (2011).

tween 2005 and 2017. We maintain a daily minimum-variance allocation based on estimates that are updated every 30 days. We find that the conditionally sparse inverse covariance matrix estimated by the graphical lasso reduces the out-of-sample volatility more than competing methods. Varying the estimation window and the number of assets we also find that these reductions are statistically significant across almost all settings. Our method many times produces portfolio volatilities that are around ten percent lower than the second best performing method. Combining performance measures into an accumulated investor utility we find that our method tends to be favored by investors with high risk aversion.

The rest of the paper is outlined as follows. Section 2 introduces the financial framework and motivates our approach, in particular the contrast of covariance and precision sparsity. Section 3 presents the graphical lasso along with a set of closely related estimators. In Section 4 we evaluate our method empirically on the S&P500. Last, Section 5 concludes our findings.

2 Minimum variance allocation and sparsity

2.1 Basic allocation and factor model framework

The global minimum variance portfolio of N risky assets is given by the solution to the following optimization,

$$\min_{w} \; w'\Sigma w \quad \text{subject to} \quad \iota' w = 1. \tag{1}$$

Here Σ denotes the $N \times N$ covariance matrix of returns, ι is a conforming vector of ones, and w is an N-vector of portfolio weights. The asset allocation that solves this minimization is

$$w = \frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota}. \tag{2}$$

It is well known that the estimation uncertainty of the sample covariance matrix can be very large. To approach this issue it is sensible to introduce a biased estimator with a lower variance to improve the out-of-sample mean-squared-error. Sparsity restrictions typically impose that a significant fraction of elements in the target matrix are zero. However, the implications of sparsity assumptions in Σ (covariance sparsity) greatly differ from restrictions in Σ^{-1} (precision sparsity). The difference between these restrictions become explicit

if we express Σ^{-1} as a function of N linear regressions of return y_i onto all other returns y_{-i} in the portfolio such that $y_i = a + \phi_i y_{-i} + u_i$. With this system of N linear regressions we can represent the inverse covariance matrix as,

$$\Sigma^{-1} = \Theta = \begin{pmatrix} \theta_{1,1}^2 & \theta_{1,2} & \dots & \theta_{1,p} \\ \theta_{2,1} & \theta_{2,2}^2 & \dots & \theta_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{p,1} & \theta_{p,2} & \dots & \theta_{p,p}^2 \end{pmatrix}. \tag{3}$$

The off-diagonal elements of the inverse are equal to $\theta_{i,j} = -\phi_{i,j}/\operatorname{var}(u_i)$ and diagonals $\theta_{i,i}^2 = \operatorname{var}(u_i)^{-1}$. Thus, precision sparsity restricts the partial linear relations between returns whereas covariance sparsity restricts pair-wise linear relations. The degree to which these assumptions are well suited for portfolio optimization depends on the financial structure of the asset returns.

An empirically relevant structure is when N returns are represented by a factor model of $K \ll N$ common factors,

$$y = \alpha + BF + e, \tag{4}$$

where F is a K-vector of factors, y is a N-vector of returns and e is a N-vector of error terms. The asset exposures B is a matrix of dimensions $N \times K$. Idiosyncratic errors have mean zero and a covariance matrix Δ, the factor covariance matrix is Ψ. Assuming that factors are non-traded, the minimum variance allocation requires the marginal inverse of returns.[2] By the Woodbury inversion lemma this is,

$$\Sigma^{-1} = (B\Psi B' + \Delta)^{-1} = \Delta^{-1} - \Delta^{-1}B(B'\Delta^{-1}B + \Psi^{-1})^{-1}B'\Delta^{-1}. \tag{5}$$

The inverse covariance matrix of returns is the difference between the idiosyncratic inverse and the effect of the marginalization, summarized in the latter term. Since neither Σ nor Σ^{-1} is sparse the relevant targets are Δ and Δ^{-1} under this structure.[3] Furthermore,

[2] While the Fama-French factors are traded portfolios we do not consider them as viable investment options.

[3] Denoting covariances between factors and returns Ξ, we partition the joint covariance matrix of factors and returns Ω,

$$\Omega = \begin{pmatrix} \Psi & \Xi \\ \Xi' & \Sigma \end{pmatrix} = \begin{pmatrix} \Psi & B\Psi \\ \Psi B' & B\Psi B' + \Delta \end{pmatrix}. \tag{6}$$

for the small-K factor structure, only Δ and Δ^{-1} pose non-standard empirical challenges. Estimating B only requires low-dimensional linear regression for each individual asset and Ψ is low-dimensional. Under a strict factor structure the idiosyncratic covariance matrix and its precision matrix are both diagonal. Hence the sparsity restrictions are trivially equivalent. However, when factor models are misspecified the loss under covariance sparsity and precision sparsity differ.

2.2 Covariance sparsity and Precision sparsity

Whether the loss under covariance sparsity is less or greater than under precision sparsity depends on the financial structure of returns. There are obvious structures that imply equivalence between covariance sparsity and precision sparsity, *e.g.* independent normally distributed returns. Other forms, such as a Toeplitz structure among covariances gives a full covariance matrix yet sparse precision matrix.[4] However, these structures imply an inherent ordering of the assets, which is economically restrictive.

An economically appealing framework is equicorrelational (EC) residual returns, *i.e.* conditional on some factors there is an omitted structure that is well described by a full correlation matrix of equal pairwise relations. In terms of the common factor structure this follows when individual residual returns $\tilde{y}_i$ are well described by equal loadings and equal error variance such that $\tilde{y}_i = fb_i + u_i \approx fb + u_i$, where b is a scalar and u_i have variance σ_u^2. Without loss of generality we consider the framework with only one common factor and no intercepts. Let σ_f^2 denote the factor variance and $\beta = b\iota$. The covariance matrix is then,

$$\tilde{\Sigma} = \sigma_f^2 \beta\beta' + \sigma_u^2 I. \tag{8}$$

Assuming uncorrelated error terms, all off-diagonal elements stem from the first term. The

Instead we approach factors and returns jointly. Inverting the covariance matrix of factors and returns reveals that the joint inverse is sparse as long N grows faster than K,

$$\Omega^{-1} = \begin{pmatrix} \Psi^{-1} + B'\Delta^{-1}B & -B'\Delta^{-1} \\ -\Delta^{-1}B & \Delta^{-1} \end{pmatrix}. \tag{7}$$

Hence, when the number of factors is small we find that Ω^{-1} is statistically sparse. The joint inverse contains all terms needed to estimate Σ^{-1} using (5). Importantly, the partition reveals that the only high-dimensional estimation challenge is the idiosyncratic inverse. This matrix is conditionally sparse, which makes it well-suited for regularized estimation.

[4] In Appendix G we demonstrate that this also holds for grouped Toeplitz structures.

relation between these parameters and the precision matrix is slightly more elusive. For the above financial model the inverse is,

$$\tilde{\Sigma}^{-1} = \Delta^{-1} - \frac{\sigma_f^2 \beta\beta'}{\sigma_u^2(\sigma_u^2 + \sigma_f^2 N b^2)} = \Delta^{-1} - \sigma_f^2 \beta\beta' c \tag{9}$$

Whereas covariances are linear in σ_f^2 the off-diagonal elements of the precision matrix are scaled by c, which is non-linear in both N and the error variance. The scaling tends to zero in N at a decreasing rate. Conversely, the scaling tends to zero as the residual variance increase for a given factor variance. Intuitively, as σ_u^2 increase, the off-diagonal elements of the inverse covariance matrix tend to zero; the partial linear relations tend to zero as the factor structure becomes negligible.

This difference gives clear implications of sparsity restrictions on portfolio volatility. Let U denote a matrix of zeroes and ones that indicates which elements are set to zero such that the element-wise product of U and Σ is sparse yet positive semi-definite. The relative portfolio volatility of a covariance sparse minimum variance allocation compared to the precision sparse alternative is then,

$$\begin{aligned} RP_w = \sqrt{\frac{w'_{(U\circ\tilde{\Sigma})}\tilde{\Sigma} w_{(U\circ\tilde{\Sigma})}}{w'_{(U\circ\tilde{\Sigma}^{-1})}\tilde{\Sigma} w_{(U\circ\tilde{\Sigma}^{-1})}}} &= \sqrt{\frac{w'_{(U\circ\Sigma)}(\sigma_f^2(b\iota)(b\iota)' + \sigma_u^2 I) w_{(U\circ\Sigma)}}{w'_{(U\circ\Sigma^{-1})}(\sigma_f^2(b\iota)(b\iota)' + \sigma_u^2 I) w_{(U\circ\Sigma^{-1})}}} \\ &= \sqrt{\frac{\sigma_f^2 b^2 + \sigma_u^2 ||w_{(U\circ\tilde{\Sigma})}||^2}{\sigma_f^2 b^2 + \sigma_u^2 ||w_{(U\circ\tilde{\Sigma}^{-1})}||^2}} \end{aligned} \tag{10}$$

The portfolio volatility under covariance sparsity exceeds that of the precision sparse allocation if $||w_{(U\circ\tilde{\Sigma})}||$ is greater than $||w_{(U\circ\tilde{\Sigma}^{-1})}||$.[5]

Let the number of non-zero elements in the respective matrices be denoted S, these elements retain the value from the common factor structure. The second norm of the minimum variance weights under precision sparsity and covariance sparsity are $||w_{(U\circ\tilde{\Sigma})}||$

[5]This also reveals that the optimal allocation under this framework is an equal weighted portfolio.

and $||w_{(U\circ\tilde{\Sigma}^{-1})}||$ respectively and the ratio of the two is,

$$R_w = \frac{||w_{\tilde{\Sigma}_s}||}{||w_{\tilde{\Sigma}_s^{-1}}||} = \frac{||\iota'(U\circ\tilde{\Sigma})^{-1}||}{||\iota'(U\circ\tilde{\Sigma}^{-1})||}\frac{|\iota'(U\circ\tilde{\Sigma}^{-1})\iota|}{|\iota'(U\circ\tilde{\Sigma})^{-1}\iota|}. \tag{11}$$

The term $||w_{(U\circ\tilde{\Sigma})^{-1}}||$ depends on the sparsity pattern imposed by U. Under sparsity restrictions that induce, for example, banded diagonal we attain equal weights under both covariance and precision sparsity. Among the patterns that imply non-equal weights $||w_{(U\circ\tilde{\Sigma}^{-1})}||$ is minimized by $U = I + (\tilde{\iota}\tilde{\iota}' - \text{diag}(\tilde{\iota}\tilde{\iota}'))$ where $\tilde{\iota}$ is an N-vector of $N - S$ ones and the remaining elements zero.[6] Given those sparsity restrictions we can approach $||w_{(U\circ\tilde{\Sigma})^{-1}}||$ as a simple partitioned inversion. Equations (8) and (9) gives the respective allocations along with the relative norm,

$$RN_w = \left(\frac{(N-S)(\sigma_u^2+\sigma_f^2b^2S)^2 + S(\sigma_u^2+\sigma_f^2b^2)^2}{(N-S)(\sigma_u^2+\sigma_f^2b^2(N-1))^2 + S(\sigma_u^2+\sigma_f^2b^2(N-S))^2}\right)^{\frac{1}{2}}\left(\frac{\sigma_u^2+\sigma_f^2b^2N^*}{\sigma_u^2+\sigma_f^2b^2S^*}\right) \tag{12}$$

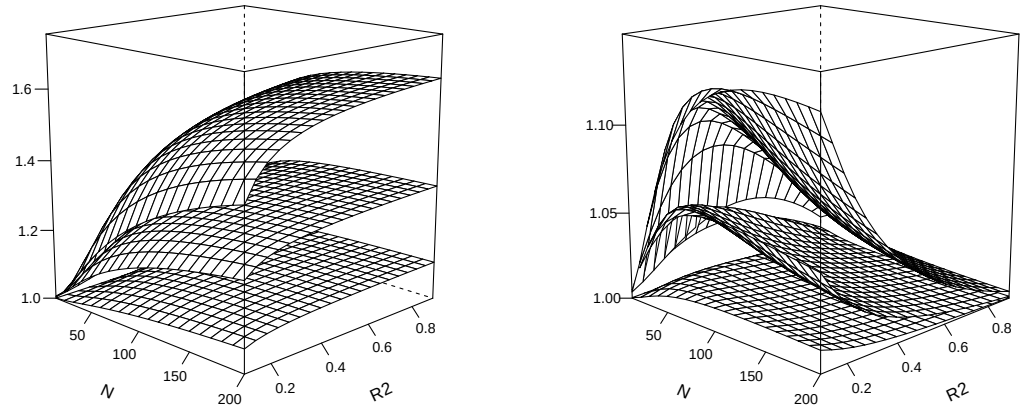
where $S^* = (S - S/N(S-1))$ and $N^* = (N - 1 - S/N(S-1))$. The ratio can be shown to be greater, or equal, to one for S smaller than N. Thus, under the strict EC structure the precision sparse restrictions reduce portfolio volatility to a greater extent than covariance sparse restrictions. The degree to which the performance differ depend on the degree of sparsity S/N as well as the ratio of σ_u^2 to $\sigma_f^2b^2$.

Figure (1) clarifies the implications of this (12), along with the relative portfolio volatility (10), as a function of the number of assets and the fit of the factor structure. The number of assets range from 10 to 200 and the univariate fit of the factor structure, $R_i^2 = R^2 = 1 - \sigma_u^2/(\sigma_f^2b^2 + \sigma_u^2)$, from 0.1 to 0.9. For convenience we restrict the figure to three levels of sparsity (S/N), twenty percent, fifty percent and eighty percent.

Differences are negligible for small N and low R^2. However, as R^2 tends to one and as N grows large, the norm of the portfolio weights under covariance sparsity exceeds the norm under precision sparse restrictions. Under less sparse representations where eighty percent of elements are non-zero the difference is greater. The norm is 60 percent greater as both N and R^2 grows large. As we turn to consider portfolio volatility, the difference is largely offset as R^2 approach 1. However, for factor structures with large error variance

[6] For details see Appendix A

Figure 1: RP_w and RN_w as a function of N and R^2



(a) Relative portfolio norm (RN_w) (b) Relative portfolio volatility (RP_w)

The ratio of the relative norm and volatility of a covariance sparse minimum variance allocation to the error of a precision sparse portfolio allocation in accordance with (12) and (10). The outer surface denote the ratios when sparsity degree (S/N) is 0.8, the inner denote results for 0.5 and 0.2 respectively. The ratio is expressed as a function of the number of assets N and the fit of the linear factor model R^2.

we see striking differences in performance between covariance sparse and precision sparse portfolios.

To gather some intuitive appreciation for the difference between precision sparse and covariance sparse restrictions it is instructive to consider the relative error when restricting the matrices to their respective diagonals. Given the EC structure the norm of the variances grow linearly in N, σ_e^2 and σ_f^2. Similarly the norm of covariances grow linearly in σ_f^2 but quadratically in N. In all, as N grows and the ratio σ_e^2/σ_f^2 tends to zero, the relative error $RN_{\tilde{\Sigma}} = ||\tilde{\Sigma} - \text{diag}(\tilde{\Sigma})||_F / ||\tilde{\Sigma}||_F$ tends to 1. Hence the relative error of the sparsity assumption increase in N and decrease in σ_e^2/σ_f^2. This follows intuitively since the commonality becomes negligible as the variance of the error term increase. Conversely, the off-diagonal elements of the precision matrix dissipates in the number of assets. As the cross sectional dimension of assets grows the respective elements tends to zero at a decreasing rate,

$$\frac{\partial c}{\partial N} = -\frac{\sigma_f^2 b^2}{\sigma_e^2(\sigma_e^2 + \sigma_f^2 N b^2)^2}. \tag{13}$$

The implications are clear in the parameters of the financial factor structure, the ratio

of the two relative errors is,

$$\frac{RN_{\tilde{\Sigma}}}{RN_{\tilde{\Sigma}^{-1}}} = \frac{||\tilde{\Sigma} - \text{diag}(\tilde{\Sigma})||_F}{||\tilde{\Sigma}^{-1} - \text{diag}(\tilde{\Sigma}^{-1})||_F} = \frac{(\sigma_f^2 b^2)^2(N-1) + (\sigma_e^2 + \sigma_f^2 b^2(N-1))^2}{(\sigma_f^2 b^2)^2(N-1) + (\sigma_e^2 + \sigma_f^2 b^2)^2}. \tag{14}$$

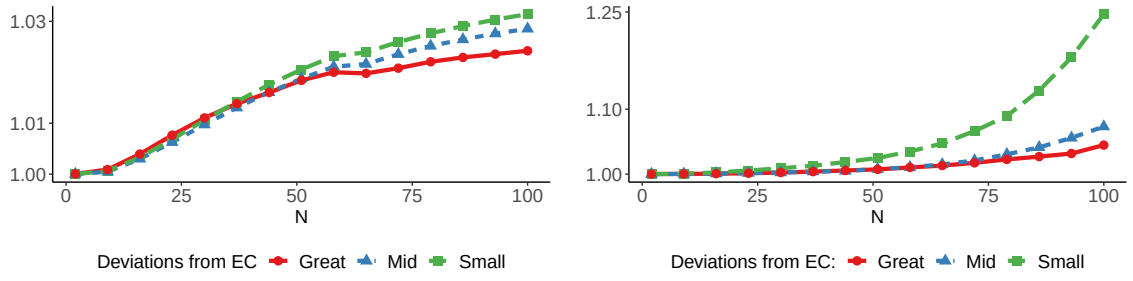
This ratio is greater than one for $N > 1$ and all σ_e^2, which underscores the previous point; the relative loss under covariance sparsity exceeds or equals the loss induced by precision sparsity.

The strict EC model favours the precision sparse restrictions, yet the picture becomes more nuanced as we allow for deviations exact framework. Relaxing the assumption that all asset exposures are equal gives different implications than relaxing the assumption of all error variances equal. The exact effect on the relative performance depends on the distributional assumptions of the respective parameters. The distinction can be illustrated by simulating deviations from the EC model. We consider two settings; (1) exposures vary but error variances are equal, (2) exposures are equal but error variances vary. For the first case we draw asset exposures b_i from a uniform distribution with mean one but range of either 0.85, 1.7 or 2.55. With all asset error variances are equal and $S/N = 0.5$ we measure the relative portfolio volatility as N increase. The draws are repeated 1000 times measuring the median ratio of covariance sparse to precision sparse volatility.[7] Likewise, for the second setting we simulate uniformly distributed error variances of increasing range. The small deviations are drawn from a range of 0.75, while mid and great deviations are drawn from 3.00 and 6.75 respectively. Since the distinction between the covariance and precision sparse restrictions is greatest when the error variance is large relative to the factor implied variance, we let the average σ_u^2 be twenty times greater than $\sigma_f^2 b^2$. Elements below the 50 percentile are set equal zero. Results of the simulations are displayed in Figure (2).

Figure 2 presents the relative portfolio volatility as we enable deviations from the EC framework. As we relax the framework by allowing variation across error variances we find that the relative gain of precision sparse restriction increase in N but at a non constant rate. For small N the differences are largely negligible, however, as N grows large the differences become more pronounced. The gains diminish as the deviations grow larger.

[7]Draws that, given the sparsity restrictions, fail to ensure that the covariance matrix is positive semi-definite are discarded. The ranges of deviations from the EC framework are calibrated such that the numerical imprecision from the covariance matrix inversion should be minimal.

Figure 2: Relative portfolio volatility under deviations from EC framework.



(a) Non-constant error variance σ_u^2 (b) Non-constant asset exposure β

The ratio of the relative error of an covariance sparse minimum variance allocation to the error of a precision sparse portfolio allocation in accordance with (12). The ratio is expressed as a function of the number of assets N for two cases of deviations from the EC covariance structure; (a) error variance varies uniformly across assets, (b) β varies uniformly across assets.

Similarly, as we enable deviations from EC by variation in exposures, the relative loss from covariance sparse restrictions become smaller as the deviations grow greater. Also, the difference between the restrictions is limited for small N. Yet there are performance gains from precision sparse representations when N is large. For small deviations the gains are substantial and growing fast in N.

In sum, we show that under an EC structure precision sparsity constraints are less restrictive than covariance sparsity constraints. The relative performance gain is greatest in the presence of weak commonalities. This results underscores the empirical relevance for portfolio applications where strong observed factors are pre-conditioned. Furthermore, as we allow for deviations from EC the gains increase in the number of assets in the portfolio, which highlights the relevance for portfolios consisting of many assets.

3 Estimating the idiosyncratic inverse

Decomposing return variation under a common factor structure lets us estimate the inverse residual covariance matrix. Targeting this inverse directly circumvents the need for any explicit large-N matrix inversion. To make this point explicit in our notation we refer to the idiosyncratic inverse as Θ. In this section we present one of the most popular methods for solving the optimization, the Glasso algorithm (Friedman et al., 2008). We also present an alternative method not based on the likelihood (15), which estimates the inverse by performing N *nodewise* regressions along the representation given in (3). This is known as

the *neighborhood* regularization method (Meinshausen and Bühlmann, 2006). This method does not impose properties such as symmetry or positive semi-definiteness on the estimate.

3.1 Inverse estimation using Graphical lasso

We estimate the idiosyncratic inverse, conditional on the financial factors, of the assets from (7). To estimate the inverse we use the negative log-likelihood, augmented by a penalization on the sum of the absolute elements. The regularization is here imposed on the first norm of the off-diagonal elements. This estimator was proposed by Yuan and Lin (2007), let $\Theta \succeq 0$ denote the set of positive semidefinite matrices,

$$\widehat{\Theta} = \text{argmin}_{\Theta \succeq 0} \left(\text{tr}(\widehat{S}\Theta) - \log(\det(\Theta)) + \Lambda ||\Theta||_1 \right), \tag{15}$$

where $\widehat{S} = T^{-1}\widehat{E}'\widehat{E}$, the covariance matrix of the residuals from (4) conditional on a set of factors. The estimate $\widehat{\Theta}$ along with the asset exposures $\widehat{B}$ and $\widehat{\Psi}^{-1}$ gives us an estimate for the marginal inverse $\widehat{\Sigma}^{-1}$ through (5). The factor covariance matrix, Ψ, is of low dimension and estimated by the sample covariance matrix does not cause any inversion problems.

We initialize the algorithm by setting the matrix S equal to the sample covariance matrix of the residuals of the Fama-French factor model. Estimating Θ we supply a strictly positive penalization parameter grid, Λ_n.[8] Under each value of the penalization grid we estimate an inverse. To find the optimal level of penalization we rely on a 5-fold cross-validation scheme. Through cross-validation we are able to directly search for the estimate that provides the lowest portfolio risk. Hence, we implement the cross-validation to find the allocation that implies the lowest portfolio volatility across the 5-folds. Randomly assigning observations into one of the five folds such that the folds remain largely equal in size, the procedure uses four fold as the estimation window to estimate the inverse and evaluates the estimates against the observations in the omitted fifth fold. Within each estimation window we estimate the inverse over a grid of regularization. We choose the level of regularization, which produces the lowest average portfolio volatility across all five validation folds. Yuan and Lin (2007) first study the properties of the estimator when the

[8] We follow common practice and select the highest level of regularization equal to $\lambda_{max} = \max\Big(\max\big(S - \mathbf{I}\big), -\min\big(S - \mathbf{I}\big)\Big)$ and the smallest level $\lambda_{min} = 0.1\lambda_{max}$ The grid between these boundaries is composed of 50 values linearly decreasing in log scale.

number of variables is fixed and the sample size increases. Rothman et al. (2008) found that for repeatedly observed *i.i.d.* multivariate normal variables the graphical lasso is consistent in the Frobenius norm under sparsity conditions and assuming that the minimum and maximum eigenvalues of the covariance matrix are bounded. They further show that the estimator is consistent at the rate $O_p(\sqrt{(N+s)\log(N)/T})$, where s denotes the number of non-zero elements of the inverse covariance matrix. Ravikumar et al. (2011) further find that these statistical properties hold for a large class of distributions beyond the multivariate Gaussian. They show that under irrepresentability conditions the estimator is consistent and they derive the decrease in the maximum norm of the element-wise error.[9,10] Under the same assumptions they also prove model selection consistency, the estimator uncovers the true cross-variable relations with the correct sign. In all, for the context at hand these papers further underscore the importance of accounting for the mutual factor structure among returns.

With the estimator calibrated, either through cross-validation or the information criteria, we estimate $\widehat{\Theta}$ over the whole estimation window. From a financial point of view this estimate describes the cross-asset hedging relations that complements the factor models. Using (5) we combine this estimate with the common factors to estimate the marginal inverse, $\widehat{\Sigma}^{-1}$. Using this estimate we form a minimum variance allocation along the lines of (2). We will hence refer to the portfolio based on this allocation as the *Factor-Graphical* portfolio.

With regards to the financial factors we consider two alternatives. A fixed small-K set up where factors are imposed *a priori*. Here factor loading and the covariance matrix is estimated using the sample covariance matrix and the least squares estimator. For this set up we use the five Fama-French factors. As an alternative to this we consider the large-K set up where we assume that a parsimonious set of relevant factors can be found among a large set of candidates. This framework introduces an additional dimensionality challenge. We approach this challenge under analogous sparsity restrictions as we impose on the hedging portfolios. To select the relevant factors we estimate the factor loading using

[9]Similar, but not equivalent to irrepresentability conditions associated with the lasso.

[10]From a graphical perspective the condition is that non-edges in the graph may not be highly correlated with the true edge set. In the phrasing of our application, spurious hedging relations may not be highly correlated with relevant hedging relations.

a lasso estimator,

$$\widehat{B}_i = \text{argmin}\left(\left|\left|y_i - a_i - B_i F\right|\right| + \lambda_i \left|\left|B_i\right|\right|_1\right). \tag{16}$$

We estimate the factor exposures for each asset independently. We estimate the precision matrix, Ψ^{-1}, for the active set of factors using the same regularized methods used to estimate Δ^{-1}. Combining the estimates of our two stages using (5) provides an estimate of Σ^{-1}.

3.2 Inverse estimation using Neighborhood Selection

There are several approaches to graphical model selection. An alternative to the graphical lasso is *neighborhood selection* (Meinshausen and Bühlmann, 2006). These methods estimate the inverse covariance matrix through a large number of regression equations. Compared to the graphical lasso the neighborhood selection methods are computationally faster and offers some additional flexibility in how we regularize the estimate. Drawback are however that these methods do not guarantee that the estimate holds properties such as symmetry and positive semi-definiteness.

Callot et al. (2019) form minimum variance portfolios using an estimator that is related to the graphical lasso. Their method broadly known as a neighborhood selection algorithm and may be written as a special case of the graphical lasso. Neighborhood selection estimators rely on different statistical assumptions and it is not clear *a priori* which is more restrictive with respect to portfolio allocation.

Let e_i denote a column of the residual covariance matrix $T^{-1}\widehat{E}'\widehat{E}$. The neighborhood selection method is an estimator based on a set of N node-wise regressions, $e_i = e_{-i}\phi_{-i} + \varepsilon_i$, where ϕ_{-i} is a parameter vector of $(N-1)$ elements. Since the number of parameters to estimate far exceeds the historical return they implement a regularized estimator,

$$\hat{\phi}_{-i}^{(\alpha)} = \text{argmin}\left(||e_i - e_{-i}\phi_{-i}|| \; + \; \tilde{\Lambda}_i ||\phi_{-i}||_\alpha\right) \tag{17}$$

where $\hat{\phi}_{-i}^{(\alpha)} = \{\hat{\phi}_{i,p};\ p = 1 \ldots N,\ p \neq i\}$. For our purposes we implement this approach under two alternative penalization terms, (1) the least absolute shrinkage selection operator estimator (lasso) and (2) the ridge estimator. The ridge estimator regularizes the estimate

through shrinkage without promoting sparsity in the inverse while the lasso also promotes sparsity. Setting the parameter α equal to one puts the penalization on the absolute parameter vector, this corresponds to a lasso estimator. Alternatively we put the penalization onto the ℓ_2 norm of the parameter vector, which provides the ridge estimator. This approach is a straightforward implementation of hedging portfolio estimation as discussed by Stevens (1998). The lasso will promote portfolios of few assets while the ridge estimator will attenuate the portfolio weights but keep hedging portfolios large.

Using either estimator we use the point estimates of (17) to construct the inverse covariance matrix. The regression of the i:th asset return onto all other returns identifies the i:th row of the inverse covariance matrix, $\widehat{\Theta}$.

$$\widehat{\Theta} = \begin{pmatrix} \hat{\theta}_{1,1}^2 & \hat{\theta}_{1,2} & \dots & \hat{\theta}_{1,p} \\ \hat{\theta}_{2,1} & \hat{\theta}_{2,2}^2 & \dots & \hat{\theta}_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\theta}_{p,1} & \hat{\theta}_{p,2} & \dots & \hat{\theta}_{p,p}^2 \end{pmatrix} \tag{18}$$

Let $\hat{\varepsilon}_i$ denote the residual of the i:th regression under the parameter vector $\hat{\phi}_{-i}^{(\alpha)}$. Using these residuals we find the elements of the inverted covariance matrix as $\hat{\theta}_{i,i}^2 = \operatorname{var}(\hat{\varepsilon}_i)^{-1}$ and $\hat{\theta}_{i,j} = -\hat{\phi}_{i,j}^{(\alpha)} / \operatorname{var}(\hat{\varepsilon}_i)$. Here the non-zero entries of the i:th row are said to be in the neighborhood of the i:th asset. By the sparsity imposed by the node-wise lasso estimates $\hat{\phi}_{-i}^{(\alpha)}$ we also impose sparsity on the estimated inverse. Sparsity is here defined on the cardinality of the parameter vectors $S_j := \{j; \phi_{-i} \neq 0\}$. In this context implies that each return is associated with a small neighborhood of assets, where small is defined in relation to the time dimension T. We calibrate the regularization parameter through K-fold cross-validation. Specifically we use a five fold cross-validation of the nodewise predictive performance to calibrate the regularization.

Meinshausen and Bühlmann (2006) also provide the theoretical properties of this approach when applied to the lasso regularization. They find that under normally distributed *i.i.d.* variables the nodewise lasso estimator is consistent for model selection under beta-min and irrepresentability conditions (Zhao and Yu, 2006).[11] The irrepresentability conditions

[11] The beta-min condition implies that the linear relation between the weakest relevant predictor is bound from below.

established by Zhao and Yu (2006) for the univariate regression setting restrict the covariation between relevant and irrelevant predictors. For the type of neighborhood selection we consider these conditions applies for all N nodewise regressions. Hence, the restriction applies to N covariance matrices of dimensions $(N-1)\times(N-1)$. Just like the Graphical estimator, Meinshausen and Bühlmann (2006) find that the initial Gaussian assumption does not appear to be crucial but can be replaced by more general conditions on the tail behavior of the variables. Wainwright (2009) for example finds that the lasso remains consistent, with some conditions on the sample growth rate, under sub-Gaussian error terms.

The neighborhood selection estimation approach, under all regularizations, suffer from drawbacks. None of these estimators impose expected properties such as symmetry and positive semi-definiteness on the estimate. To ensure that the resulting empirical concentration matrix is positive semidefinite we need to rely on eigenvalue cleaning procedures. Hautsch et al. (2012) propose a cleaning procedure based on separating the noise-driven eigenvalues from those arising due to signal in the financial markets. The latter eigenvalues remain unaffected while noise-driven values are adjusted away from zero. Let v_i denote an eigenvalue and V_i the corresponding eigenvector. We apply a procedure that adjusts all negative eigenvalues to equal to the smallest positive value,[12]

$$\hat{v}_i = \begin{cases} v_i & \text{if } v_i > 0 \\ \bar{v}, & \text{otherwise} \end{cases} \tag{19}$$

where $\bar{v} = \min\Big(v_i \big| (v_i > 0)\Big)$. Using the vector of cleaned eigenvalues, $\hat{v}$, along with corresponding eigenvectors we can project a semidefinite inverse, $\widetilde{\Theta} = V\,\mathrm{diag}(\hat{v})V^{-1}$. Using this estimate we may again proceed with the portfolio allocation under the minimum variance allocation (2). In our empirical evaluation we refer to the portfolios allocated under these estimators as *Nodewise* ℓ_1 and *Nodewise* ℓ_2.

[12]Since there is no need to invert this matrix we do not need to worry about possible numerical issues associated with small eigenvalues.

4 Minimum variance allocation of the S&P500

In this section we evaluate the performance of the Factor-Graphical method of portfolio allocation using S&P500 companies. We first introduce our data and the financial performance measures we use to assess performance. The main results that follow compares our financial performance against the main alternative methods: Shrinkage (Ledoit and Wolf, 2004), Non-linear Shrinkage (Ledoit and Wolf, 2015), POET (Fan et al., 2013), Neighborhood selection and $1/N$. Full details of how we implement the Shrinkage and POET approaches are presented in Appendix E. The $1/N$ rule allocates the total wealth equally over all assets. While seemingly naive this allocation rule has often proved difficult to outperform empirically (DeMiguel et al., 2007). Following our main performance comparison we analyze whether our performance stems disproportionately from the factors or the hedging portfolios by contrasting our performance against two restricted versions of the method. The first is a strict factor model. The second applies a graphical lasso directly to the returns without conditioning on factors.

4.1 Data and performance measures

Daily asset returns are collected from the CRSP database spanning the years 2010 until 2018. Since the companies comprising the S&P500 vary over time we include all companies that at some point were included during the period 2005-2017. This provides a set of almost 800 assets. From this set of assets we form 6 balanced panels of N assets. We consider the sample dimensions $N = \{50, 100, 150, 300, 400, 500\}$. For each given number of assets portfolios the estimates of $\widehat{\Sigma}^{-1}$ are updated very 30 trading days. At the point of re-estimation the respective procedures are provided an estimation window of one among three sizes, $\tilde{t} = \{50, 150, 250\}$, of historical daily returns. The estimated allocation is maintained under daily re-balancing until the next re-estimation. For our basic Factor-Graphical specification we condition on the five Fama-French factors. For our large-K specification we consider the 50 anomaly characteristics used by Kozak et al. (2019), which also includes the Fama-French factors. Total number of returns are denoted T. Basic descriptive statistics and individual asset performance measures are presented in Appendix D.

We compare the performance of all the estimation methods by the out-of-sample volatility, Sharpe ratio, turnover and investor utility under a quadratic utility function including transaction costs. With allocation updates every 30 days we get $M = (T - \tilde{t})/30$ portfolio allocations, indexed $m = 1 \dots M$, each used over it respective following 30 days. The portfolio weights $\hat{\mathbf{w}}_m^P$ are estimated using observations preceding the month m. We compute the realized average daily return and portfolio volatility for each method P,

$$\hat{\mu}_P = (T - \tilde{t})^{-1} \sum_{m=1}^{M} \sum_{d=1}^{30} \hat{\mathbf{w}}_m^P r_{m,d}, \tag{20}$$

$$\hat{\sigma}_P = \sqrt{(T - \tilde{t} - 1)^{-1} \sum_{m=1}^{M} \sum_{d=1}^{30} \left(\hat{\mathbf{w}}_m^P r_{m,d} - \hat{\mu}_P \right)^2}, \tag{21}$$

where $\mathbf{w}_m^P$ are weights and $\mathbf{r}_{m,d}$ is a vector of returns at day d. With the realized returns and the portfolio volatility we compute the Sharpe ratio $SR_P = \hat{\mu}_P / \hat{\sigma}_P$. Due to transaction costs on financial markets an important performance metric is how much trade is required to maintain a portfolio allocation. Let the turnover of forming the portfolio at day d of month m be denoted $\delta_{m,d}^P$. The average turnover is,

$$\hat{\delta}_P = (T - \tilde{t})^{-1} \sum_{m=1}^{M} \sum_{d=1}^{30} \sum_{i=1}^{N} \left| \widehat{w}_{i,m}^P - \widehat{w}_{i,m}^* \frac{1 + r_{i,m,d+1}}{1 + \widehat{\mathbf{w}}_m^P \mathbf{r}_{m,d+1}} \right|, \tag{22}$$

The vector $\mathbf{r}_{m,d+1}$ denotes the returns on $d+1$ of month m and the weights $\widehat{w}_{i,m}^*$ are defined,

$$\widehat{w}_{i,m}^* = \begin{cases} \widehat{w}_{i,m-1}^P, & \text{if } d = 1 \\ \widehat{w}_{i,m}^P, & \text{if } d > 1 \end{cases} \tag{23}$$

We present basic descriptive measures for portfolios, including weight dispersion, range and fraction of negative weights in Appendix F.

4.2 Factor-Graphical portfolio performance

This subsection presents our main results. The Factor-Graphical portfolio performance is compared to the principal methodologies of the literature. Additionally we consider the neighborhood selection estimators of the idiosyncratic inverse. We summarize the financial

performance of the respective portfolio allocation in Table 1. With daily portfolio returns we ultimately observe almost 1500 out-of-sample daily volatilities. Since this is a relatively large sample of observations we make use of the test derived in Ledoit and Wolf (2011), to test the null hypothesis that the volatilities of two return series are equal. The p-value of this test is computed under HAC standard errors.

We find that the Factor-Graphical method consistently produces very low portfolio volatility. Considering the cases where $N < \tilde{t}$, the method produce annualized volatilities as low as 9.57 percent while competing methods never produce volatilites below 10 percent. Generally, the non-linear shrinkage method of Ledoit and Wolf (2015) is the most successful alternative. However, the volatility reductions produced by the Factor-Graphical method are statistically significant at 5 percent level for all cases except for the case where both N and $\tilde{t}$ both equal 50. Further, the two Nodewise estimators perform at a level comparable to the prominent covariance estimators of the literature when the estimation window grows or the portfolio sizes increase to 100 or 150 assets.

Turning to the case where $N > \tilde{t}$ we find stark differences in performance between the methods. In these settings, the Factor-Graphical method significantly outperforms the alternatives at a 1 percent significance level for all cases except for when N is 500 and $\tilde{t}$ equals 50. The differences are particularly striking for the larger estimation windows. For example, with N equal 500 and $\tilde{t}$ is 250 the Factor-Graphical method gives a portfolio volatility of 8.16 percent, annualized, compared to the Non-linear shrinkage approach which gives 9.51 percent. We also find that the nodewise estimators provide compelling alternatives for these high-dimensional portfolios. However, surprisingly we do not find any big differences between the ℓ_1 and ℓ_2 regularization.

A somewhat surprising result is the portfolio volatility often increase slightly as we increase the estimation window from 150 to 250 observations. We speculate that the decrease in performance at larger sample windows is the result of a bias-variance trade-off. Parameter instability in the true covariance matrix along with time-varying factor exposures would render older historical returns less informative than recent. The optimal sample window size considers both the rate of decay in the informativeness and the rate of convergence in the estimator. This is a big empirical question, which is beyond the scope of this paper.

The portfolios allocated relying on the covariance estimates of the Linear shrinkage and

Table 1: Financial Performance 2010-03-31 until 2016-12-28.

		$\tilde{t} = 50$			$\tilde{t} = 150$			$\tilde{t} = 250$		
	M = 30	$\hat{\sigma}$	$\hat{\mu}/\hat{\sigma}$	$\hat{\delta}$	$\hat{\sigma}$	$\hat{\mu}/\hat{\sigma}$	$\hat{\delta}$	$\hat{\sigma}$	$\hat{\mu}/\hat{\sigma}$	$\hat{\delta}$
N = 50	Factor-Graphical	12.49	0.93	1.87	10.89*	1.36	1.83	10.85**	1.19	1.80
	Nodewise ℓ_1	13.48	0.80	1.71	11.96	1.28	1.37	11.62	1.38	1.30
	Nodewise ℓ_2	13.56	0.96	1.76	11.91	1.14	1.36	11.44	1.47	1.32
	Shrinkage	13.24	1.05	2.34	11.60	1.46	2.39	11.39	1.18	2.17
	NL-Shrinkage	12.77	1.11	2.02	11.14	1.50	2.14	11.15	1.23	2.04
	POET	12.49	0.92	1.61	11.70	1.23	1.75	12.23	1.09	1.78
N = 100	Factor-Graphical	11.82**	0.86	2.31	9.97**	1.46	2.41	10.03**	1.45	2.41
	Nodewise ℓ_1	12.66	0.87	2.49	10.98	1.22	1.52	11.10	1.20	1.52
	Nodewise ℓ_2	12.59	0.86	2.49	11.09	1.28	1.52	11.01	1.24	1.49
	Shrinkage	12.86	0.77	2.98	11.29	1.32	3.86	11.13	1.32	3.67
	NL-Shrinkage	12.38	0.83	2.37	10.70	1.45	3.30	10.68	1.38	3.00
	POET	13.32	0.71	1.79	13.04	1.05	2.05	14.15	1.01	2.15
N = 150	Factor-Graphical	11.41**	0.96	2.78	9.57**	1.74	2.97	9.80**	1.43	2.99
	Nodewise ℓ_1	11.82	1.00	2.93	11.00	1.30	1.87	10.80	1.34	1.76
	Nodewise ℓ_2	11.84	1.05	2.91	11.18	1.15	1.86	11.06	1.32	1.76
	Shrinkage	12.45	0.98	3.34	11.09	1.72	5.04	11.26	1.45	5.08
	NL-Shrinkage	12.10	1.00	2.63	10.15	1.86	3.41	10.66	1.38	3.92
	POET	12.91	0.76	1.85	12.88	1.15	2.06	13.67	1.10	2.12
N = 300	Factor-Graphical	10.92**	1.24	3.21	8.78**	1.94	3.65	9.16**	1.82	3.82
	Nodewise ℓ_1	11.25	1.22	3.23	9.89	1.67	2.42	10.13	1.46	2.12
	Nodewise ℓ_2	11.23	1.22	3.23	9.88	1.69	2.41	10.16	1.55	2.10
	Shrinkage	11.90	1.04	3.37	10.41	1.90	5.88	11.31	1.42	7.16
	NL-Shrinkage	11.70	1.02	2.84	9.59	1.93	3.88	10.11	1.70	4.16
	POET	12.25	0.82	2.01	12.22	1.26	2.30	13.14	1.15	2.40
N = 400	Factor-Graphical	10.72**	1.21	3.24	8.41**	1.94	3.86	8.77**	1.79	4.05
	Nodewise ℓ_1	11.17	1.22	3.30	9.59	1.81	2.86	9.87	1.68	2.34
	Nodewise ℓ_2	11.15	1.23	3.30	9.58	1.68	2.88	9.86	1.54	2.35
	Shrinkage	11.79	1.00	3.35	9.95	1.75	5.76	11.02	1.50	7.45
	NL-Shrinkage	11.62	0.98	2.90	9.31	1.77	4.02	9.84	1.37	4.57
	POET	12.06	0.81	2.02	11.90	1.26	2.32	12.64	1.15	2.41
N = 500	Factor-Graphical	11.24	1.15	3.32	8.10**	2.04	3.99	8.61**	1.77	4.26
	Nodewise ℓ_1	11.59	1.07	3.35	9.07	2.06	3.39	10.04	1.44	2.62
	Nodewise ℓ_2	11.59	1.07	3.35	9.28	1.82	3.39	10.02	1.57	2.62
	Shrinkage	11.65	1.21	3.32	9.44	1.59	5.62	10.42	1.44	7.30
	NL-Shrinkage	11.54	1.19	2.94	9.03	1.68	4.12	9.51	1.38	4.54
	POET	12.09	0.86	2.08	11.72	1.33	2.41	12.54	1.20	2.52

The table presents the out-of sample performance for portfolios using one of five alternative estimation approaches. The parameters N, M and $\tilde{t}$ refer to the number of assets, re-estimation frequency and the estimation window length, respectively. The volatilities $\hat{\sigma}$, turnover $\hat{\delta}$ and Sharpe ratios $\hat{\mu}/\hat{\sigma}$ are all annualized through a $\sqrt{252}$ or 252 scaling. The portfolio volatilities are scaled by 100, (*) and (**) indicate that the volatility is significantly lower than *all* competing methods at 0.05 and 0.01 percent significance levels, respectively.

POET methods, do not reduce the portfolio volatility to the same extent as the Factor-Graphical nor the Nodewise estimators does. We find that the portfolios based on the POET estimator produce volatilities varying between 11.90 and 13.14 percent. The reason for this comparably poor performance is not fully understood but can likely in part be attributed to the dual estimation challenge of having to estimate the leading eigenvalue as well as the variances of the residuals to those leading $\hat{k}$ factors. Of the covariance estimators the Non-linear shrinkage approach provides the lowest portfolio volatility in almost all cases.

As a secondary consideration to the portfolio volatility we compute the Sharpe ratios. For the Factor-Graphical method we find that the annualized Sharpe ratios vary between 0.86 and 2.04. Across the portfolio sizes we find little evidence of a single dominating method. However, generally the POET seems to provide lower Sharpe ratios than the alternatives. As we consider the especially high-dimensional cases of $N = \{300, 400, 500\}$ we mostly find that one of the inverse covariance estimators provide the highest ratio. The Linear and Non-linear shrinkage methods appear largely comparable in terms of Sharpe ratios.

As a third performance measure we consider the portfolio turnover. What is striking is the often high-turnover associated with the Linear shrinkage approach. In some cases when the number of assets grow large the annualized daily turnover of this method is above 7 percent. The Factor-Graphical portfolios are generally among the more costly allocations to maintain, but they never exceed an annualized daily turnover of 4.26 percent. The Non-linear shrinkage approach also tend to be costly to maintain. For the longer estimation windows of 150 and 250 observations we find that the nodewise estimators tend to produce the most stable portfolios. Furthermore, the portfolios based on the POET estimator also incur relatively low trading costs. This, together with the relatively high volatility of these portfolio allocations, could indicate that this method tend to impose excessive sparsity.

In sum, we have found that the method we propose in this paper appears to have strong volatility reduction properties. The Factor-Graphical method reduces risk beyond alternative inverse estimations, nodewise ridge and lasso, as well as the main covariance estimators of the literature. Also the Nodewise estimators provide very competitive volatility reductions, which seems to imply that there are significant gains to be made from circumventing the large-N matrix inversion that occur in large minimum variance port-

folios. Among the covariance estimators we find that the most competitive alternative is the Non-linear shrinkage approach. However, the volatility reductions associated with the Factor-Graphical method are statistically significant in all but a few cases. In Appendix 4.4 we bring all measures together into investor utility we find that the risk averse actors tend to prefer our Factor-Graphical, or the nodewise implementation of our method, despite the higher trading costs.

4.3 Decomposing the Factor-Graphical portfolio and the large-K challenge

The Factor-Graphical portfolio combines financial factor and sparse hedging portfolios to form the minimum variance allocation. In this section we explore how these two components emerge empirically and whether we can achieve a similar level of financial performance while restricting one of the two components. In addition we consider the large-K case of many candidate factors but no prior assumptions on which of them are relevant.

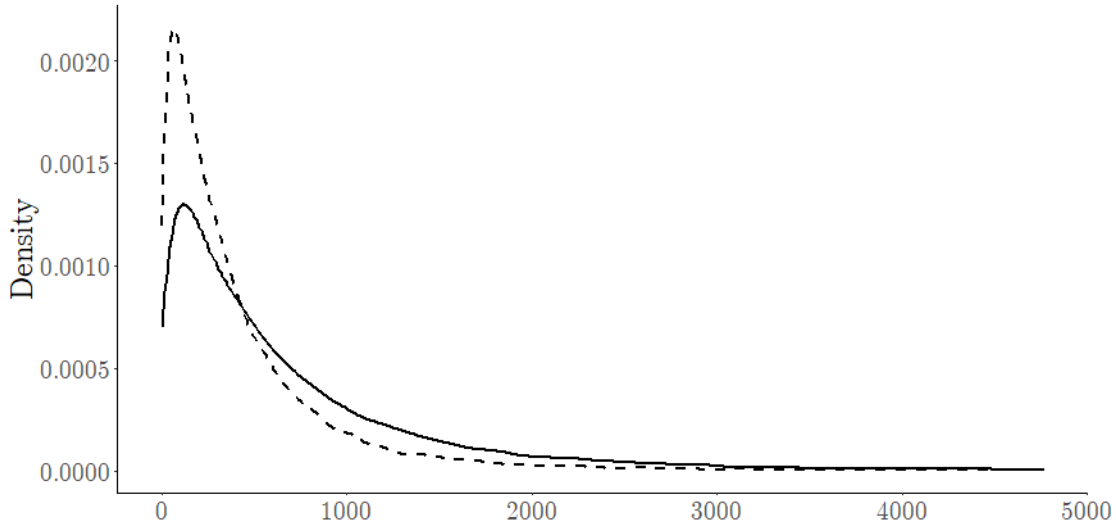
Our first evaluation is with respect to the implementation of the factors. Do the factors increase sparsity in the inverse, and how are the hedging relations developing over time. The factors must be sufficiently predictive of expected returns in order to induce sparsity in the inverse matrix. To give an indication of the factor potential, while avoiding the statistical issues of high-dimensionality, we randomly select 10 assets of the S&P500 observed over 24 months following a random date between January 1st 2010 and December 31st 2015. Estimating the idiosyncratic inverse under the Fama-French 5-factor model we find a clear attenuation of the off-diagonal elements.[13] Repeating this random selection and estimation L times provides two smooth densities for the estimates absolute off-diagonal elements.

Figure 3 displays the densities of the absolute off-diagonal elements of the two inverses. The factors induce a shift in the distribution of hedging relations, hence aiding our sparsity assumption.

The second question we address is how the level of sparsity varies over time. We estimate the inverse covariance matrix monthly with the graphical lasso. Each of the estimates provides hedging portfolios for every asset. Allowing three estimation windows of 50, 150 and 250

[13]The estimate of the inverse is based on the sample covariance matrix which works well for this purpose since the number of asset is very small.

Figure 3: Off-diagonal elements Θ and Σ^{-1}



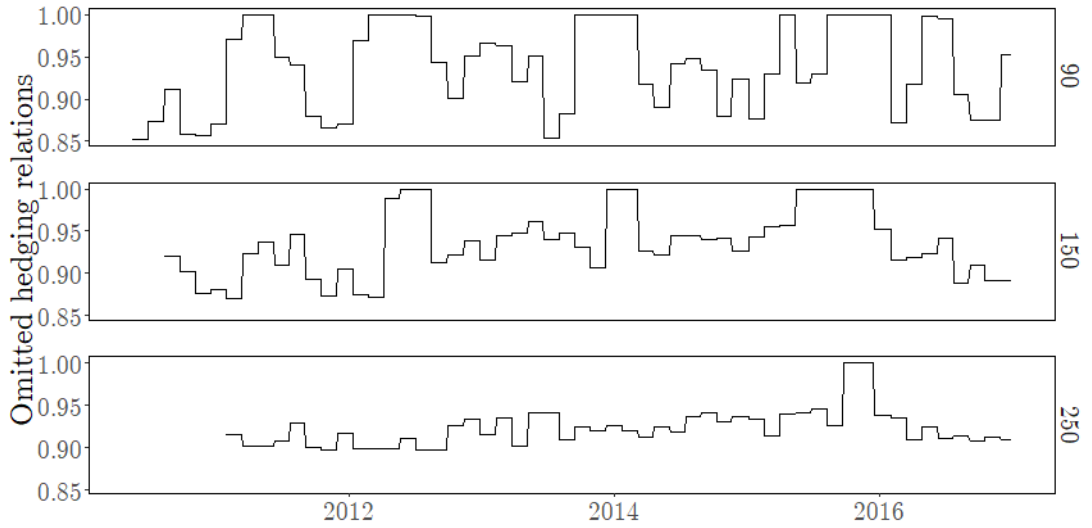
Empirical density of absolute off-diagonal elements from low-dimensional estimates of Θ and Σ^{-1}. Off-diagonal elements of idiosyncratic inverse $\widehat{\Theta}$(dotted) and off-diagonal elements of $\widehat{\Sigma}^{-1}$(solid). The two inverse covariance matrices are estimated based on historical financial returns. One thousand times, ten assets are randomly selected from the S&P500 and observed during 24 months period with a random start date between 2010-01-01 until 2015-12-31.

observations we find that the estimator tends to omit between 85 and 100 percent of the possible hedging coefficients. In the larger sample of 250 historical returns we find the most stable level of sparsity with between 5 and 10 percent of the available hedging relations used. Figure 4 displays the fraction of hedging coefficients that are set equal to zero by the graphical lasso.

If the fraction is consistently equal to one the estimator is simply reduced to a strict factor structure. However, with the largest estimation window there is only one instance in which the estimator omit all possible hedging relations. Under smaller estimation windows this is a much more frequent event, but the sparsity generally varies within the range of 85 to 100 percent. In all this implies that a strict factor structure may be too simple.

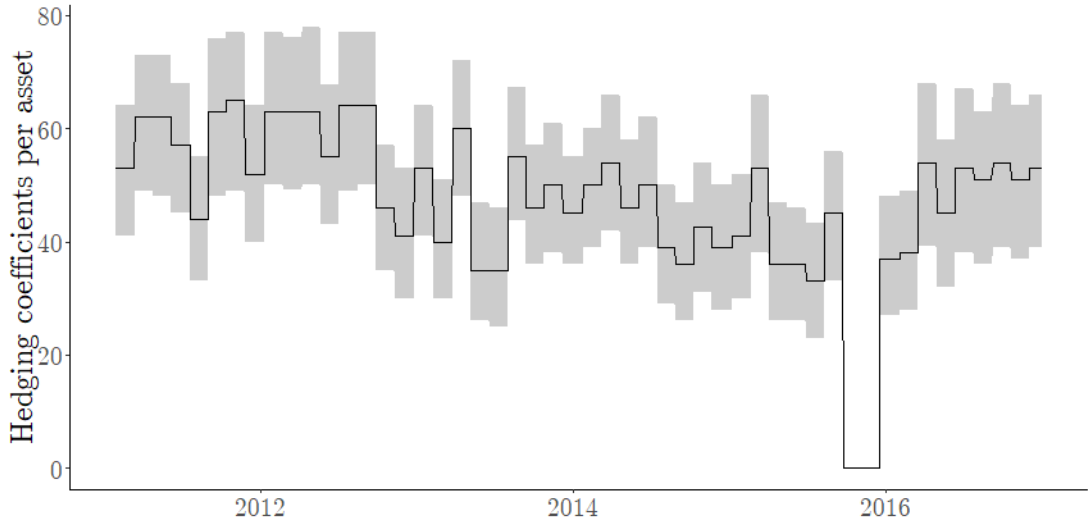
In the same way we follow the fraction of hedging relations in the whole portfolio, we also track the distribution of hedges per asset. Figure 5 display the median along with the 10th and 90th percentile of hedging relations per asset. The median number hedges per asset varies largely between 65 and 0, when the estimation window consists of 250 observations. During the last months of 2015 zero hedging relations are included; all off-diagonal elements in the idiosyncratic inverse were set to zero.

Figure 4: Realized hedging portfolio sparsity over time.



Fraction of hedging coefficients omitted by the graphical model selection under 3 sample windows, $\tilde{t} = \{50, 150, 250\}$. The total number of potential hedging relations is $N(N-1)/2$. The graphical lasso performs model selection to set the irrelevant relations equal to zero.

Figure 5: Distribution of non-zero hedging relations per asset over time.



Number of hedging relations per asset. Based on estimate using window $\tilde{t} = 250$. The solid line denotes the median number of hedging relations selected by the graphical model selection procedure. The shaded area displays the range between the 10th and 90th percentile of the number of hedging relations selected for an asset.

The shaded area of Figure 5 denotes the interval between, and including, the 10th and 90th percentile. From this plot we find that the number of assets in the hedging portfolios largely varies between 80 and 25. Apart from the anomaly of the last months of 2015 it appears that the estimated optimal number of assets necessary to hedge the residual risk is largely stable.

Next we evaluate whether variations of the Factor-Graphical allocation can achieve similar performance. We consider three alternatives to the basic approach. We evaluate whether performance stems from the combination of factors and the sparse hedging. For this we introduce two restricted alternatives to our method. First, a strict factor portfolio, where the estimate is based on a linear combination of the factor covariance matrix, scaled by the asset factor loadings, and a diagonal residual covariance matrix. Second, a sparse hedging portfolio, which is based on a graphical lasso estimate but does not partition or make use of any financial factor structure. This portfolio is allocated using the graphical lasso to directly estimate the inverse covariance matrix of financial returns; the marginal inverse covariance matrix. We refer to the first portfolio as the Factor and the second as the Graphical allocation. We also evaluate whether the more flexible factor structures can further improve the conditions for sparsity. For this we expand the set of factors K to include 50 candidates. This Large-K Factor-Graphical approach still relies on regularization but in two stages, first selecting relevant factors then selecting relevant hedges. Lastly, we include an allocation of equal weights, $1/N$, to evaluate the gains made by any Factor-Graphical variation compared to a 'naive' allocation rule.

The outcome of this comparison is summarized in Table 2. The comprehensive Factor-Graphical method achieves the lowest portfolio volatility and outperforms the restricted versions. For all allocations of 300, 400 and 500 assets we find that the volatility reductions are statistically significant at the 1 percent significance level across all alternative specifications. It also consistently produces higher Sharpe ratios. However, both the Factor and Factor-Graphical allocations tend to produce higher turnover than the other estimation methods.

Comparing the two reduced forms we find that the strict Factor structure produce lower volatility compared to the Graphical allocations when the estimation window remains small. This turn in favor of the Graphical approach when the window is increased. The

only exception to this is when the portfolio is restricted to only 50 assets. While both of these methods reduce volatility and increase Sharpe ratios to a greater extent than the $1/N$ allocation rule, neither improve performance to the same extent as the methods that combines factors and regularized hedging.

Table 2: Financial Performance 2010-03-31 until 2016-12-28.

		$\tilde{t} = 50$			$\tilde{t} = 150$			$\tilde{t} = 250$		
	$M = 30$	$\hat{\sigma}$	$\hat{\mu}/\hat{\sigma}$	$\hat{\delta}$	$\hat{\sigma}$	$\hat{\mu}/\hat{\sigma}$	$\hat{\delta}$	$\hat{\sigma}$	$\hat{\mu}/\hat{\sigma}$	$\hat{\delta}$
$N = 50$	Factor-Graphical	12.49	0.93	1.87	10.89	1.36	1.83	10.85**	1.19	1.80
	Large-K Factor-Graphical	12.75	0.62	1.53	10.98	1.05	1.64	11.05	1.11	1.57
	Factor	12.67	0.88	1.97	11.43	1.27	1.97	11.77	1.09	1.92
	Graphical	13.06	1.03	0.79	11.84	1.22	0.82	11.91	1.26	0.85
	$1/N$	16.67	0.94	0.86	16.09	1.01	0.85	16.18	0.91	0.86
$N = 100$	Factor-Graphical	11.82	0.86	2.31	9.97*	1.46	2.41	10.03**	1.45	2.41
	Large-K Factor-Graphical	12.00	0.40	1.76	10.17	0.83	2.26	10.27	1.04	2.38
	Factor	12.40	0.74	2.20	11.57	1.24	2.37	12.41	1.14	2.39
	Graphical	12.47	0.93	0.89	10.99	1.23	0.97	11.18	1.29	1.03
	$1/N$	15.40	0.94	0.84	14.84	1.03	0.83	14.93	0.93	0.83
$N = 150$	Factor-Graphical	11.41**	0.96	2.78	9.57	1.74	2.97	9.80*	1.43	2.99
	Large-K Factor-Graphical	11.83	0.44	1.91	9.62	1.20	2.70	10.00	1.13	2.98
	Factor	12.19	0.92	2.41	11.56	1.54	2.76	12.56	1.20	2.84
	Graphical	12.57	0.96	0.96	11.02	1.26	1.08	11.15	1.30	1.15
	$1/N$	16.55	0.86	0.89	15.93	0.94	0.89	16.05	0.85	0.89
$N = 300$	Factor-Graphical	10.92**	1.24	3.21	8.78**	1.94	3.65	9.16**	1.82	3.82
	Large-K Factor-Graphical	11.53	0.45	2.11	9.03	1.13	3.14	9.38	1.25	3.65
	Factor	11.71	1.05	2.62	10.96	1.68	3.26	11.86	1.30	3.47
	Graphical	12.61	1.05	1.23	10.69	1.28	1.30	11.01	1.30	1.43
	$1/N$	17.44	0.85	0.93	16.75	0.94	0.92	16.84	0.86	0.92
$N = 400$	Factor-Graphical	10.72**	1.21	3.24	8.41**	1.94	3.86	8.77**	1.79	4.05
	Large-K Factor-Graphical	11.52	0.52	2.16	8.83	1.16	3.19	9.10	1.23	3.75
	Factor	11.59	1.07	2.64	10.64	1.67	3.34	11.42	1.25	3.60
	Graphical	12.73	1.04	1.34	10.71	1.32	1.41	11.07	1.28	1.54
	$1/N$	17.68	0.84	0.97	16.97	0.92	0.96	17.06	0.83	0.96
$N = 500$	Factor-Graphical	11.24**	1.15	3.32	8.10**	2.04	3.99	8.61**	1.77	4.26
	Large-K Factor-Graphical	11.79	0.56	2.19	8.53	1.29	3.24	8.84	1.30	3.84
	Factor	11.92	1.07	2.66	10.26	1.78	3.41	11.14	1.34	3.73
	Graphical	12.89	1.03	1.39	10.73	1.41	1.48	10.97	1.41	1.65
	$1/N$	17.88	0.85	0.98	17.17	0.94	0.97	17.27	0.84	0.98

The table presents the out-of sample performance for portfolios using one of five alternative estimation approaches. The parameters N, M and $\tilde{t}$ refer to the number of assets, re-estimation frequency and the estimation window length, respectively. The volatilities $\hat{\sigma}$, turnover $\hat{\delta}$ and Sharpe ratios $\hat{\mu}/\hat{\sigma}$ are all annualized through a $\sqrt{252}$ or 252 scaling. The portfolio volatilities are scaled by 100, (*) and (**) indicate that the volatility is significantly lower than *all* competing methods at 0.05 and 0.01 percent significance levels, respectively.

We find that the Large-K setting does not produce the same level of volatility reduction as the basic Factor-Graphical method. This could be due to the additional estimation

uncertainty or the need to perform a K dimensional matrix inversion when forming the allocation. The differences are not statistically significant in all small-N allocations, but across all large-N portfolios we find significant differences. Also, Sharpe ratios tend to be lower under this allocation method, but turnover also remains lower.

The last of our main results concern a recurring question for statistically driven methods of portfolio allocation; does the method outperform an equal $1/N$ allocation across the assets. This is an especially relevant consideration as the number of assets grows large. The performance measures in Table 2 makes clear that there are significant gains in variance reduction and Sharpe performance to be gained from a more involved estimation scheme. The volatility of our Factor-Graphical allocation is only 35 percent of the volatility achieved by the $1/N$ benchmark. As for the Sharpe ratio, our method more than doubles the performance of the $1/N$ portfolio. The caveat is predictably in the occurred transaction costs. The Factor-Graphical portfolios have higher turnover and are therefore more costly to maintain.

In all it appears that the performance of the Factor-Graphical approach stems from the combination of the data-driven hedging portfolio selection together with the financial factors. Neither of the reduced form approaches provide the same level of volatility reduction. Expanding the set of factors beyond the five Fama-French factors does not provide improvements to volatility reductions, but still remains comparable when the estimation window and portfolio dimensions remain low.

4.4 The economic value of precision sparsity

To provide a measure that combines the turnover, portfolio return and the volatility, we use a utility based approach similar to Fleming et al. (2001, 2003). Assuming a quadratic utility function for an investor with risk aversion κ, we get the realized utility

$$U_{m,d}^{P} = (1 + \mathbf{w}_{m}^{P}\mathbf{r}_{m,d} - c\delta_{m,d}^{P}) - \Upsilon\Big((1 + \mathbf{w}_{m}^{P}\mathbf{r}_{m,d} - c\delta_{m,d}^{P})\Big)^{2}, \tag{24}$$

where $\Upsilon = \kappa/(2(1+\kappa))$ and κ denotes the relative risk aversion. The parameter c denotes transaction costs, which are imposed proportionally to the turnover. We follow Garlappi et al. (2007) and set the costs equal to 50 basis points. To compare the different methods we compute how large a constant daily fee must be for investors to become indifferent

between two methods. Specifically, we find the fee, χ, such that the utility gain under the Factor-Graphical method equals that of the competing method, $\sum_{m=1}^{M}\sum_{d=1}^{30} U_{m,d}^{FG}(\chi) = \sum_{m=1}^{M}\sum_{d=1}^{30} U_{m,d}^{P}$ with.

$$U_{m,d}^{FG}(\chi) = \Big((1 + \mathbf{w}_m^{FG}\mathbf{r}_{m,d} - c\delta_{m,d}^{FG} - \chi\Big) - \Upsilon\Big(1 + \mathbf{w}_m^{FG}\mathbf{r}_{m,d} - c\delta_{m,d}^{FG} - \chi\Big)^2. \tag{25}$$

Finding that χ is negative implies that the realized utility of the Factor-Graphical method is lower than the competing method. These pair-wise comparisons are carried out between the Factor-Graphical portfolios and all alternative estimators.

Table 3 reports the annualized estimated fees in percent. Positive fees implies that the Factor-Graphical portfolio holds the higher utility while a negative implies that the competing method is higher.

First we contrast the utility gains for high risk aversion investors. Under high risk aversion the Factor-Graphical approach tend to outperform the covariance estimators when the number of assets increase beyond 50. The highest implied fees are associated with the comparison between the Factor-Graphical approach and the Linear shrinkage method. This is likely due to the excessive turnover these portfolios incur. Turning to the inverse covariance estimators based on nodewise estimation we find that these estimators appear generally well on par with our main estimation approach. Again, this is likely attributable to turnover, since these allocations generally provide lower volatility reductions but, however, tend to incur less trading costs.

As for the investors with lower risk aversion, $\kappa = 1$, the Factor-Graphical method never provides the best alternative. The volatility reductions are generally not large enough to offset the additional costs. For this level of risk aversion the highest utility tend to be provided by either the Nodewise alternatives or the Non-linear shrinkage approach. Restricting the comparison to the covariance matrix estimators we see that the comparison often shift in favor of the Factor-Graphical when the portfolios grow larger and with estimation windows above 50. There are some exception to this, notably associated with the POET estimator, which due to its very low turnover sometimes provides higher utility than the Factor-Graphical method even when both portfolios and $\tilde{t}$ are large.

Estimating the fees necessary to create indifference among highly risk averse investors

Table 3: Estimated equalization fee (χ_κ) on Factor-Graphical portfolio

		$\tilde{t} = 50$			$\tilde{t} = 150$			$\tilde{t} = 250$		
	M = 30	χ_1	χ_5	χ_{10}	χ_1	χ_5	χ_{10}	χ_1	χ_5	χ_{10}
N = 50	NW ℓ_1	0.69	1.21	1.86	−0.94	−0.45	0.16	−3.60	−3.26	−2.82
	NW ℓ_2	−1.40	−0.84	−0.14	0.72	1.19	1.78	−4.38	−4.12	−3.79
	Shrinkage	−1.62	−1.23	−0.75	−1.38	−1.05	−0.65	0.01	0.24	0.54
	NL-Shrinkage	−2.37	−2.22	−2.04	−1.48	−1.37	−1.23	−0.40	−0.27	−0.10
	POET	−0.29	−0.29	−0.29	0.45	0.81	1.27	−0.28	0.36	1.16
N = 100	NW ℓ_1	−0.60	−0.19	0.33	0.14	0.57	1.10	0.18	0.63	1.20
	NW ℓ_2	−0.45	−0.07	0.40	−0.71	−0.24	0.36	−0.15	0.26	0.78
	Shrinkage	1.13	1.64	2.29	1.61	2.17	2.88	1.53	1.99	2.58
	NL-Shrinkage	−0.01	0.26	0.60	0.27	0.57	0.95	0.55	0.82	1.16
	POET	0.13	0.88	1.82	0.69	2.11	3.88	0.38	2.38	4.88
N = 150	NW ℓ_1	−0.63	−0.44	−0.20	1.15	1.74	2.48	−1.88	−1.47	−0.94
	NW ℓ_2	−1.18	−0.98	−0.73	2.55	3.22	4.06	−2.06	−1.53	−0.87
	Shrinkage	−0.39	0.10	0.72	0.32	0.95	1.74	0.47	1.09	1.87
	NL-Shrinkage	−1.24	−0.92	−0.52	−1.62	−1.39	−1.10	0.58	0.94	1.39
	POET	0.19	0.91	1.82	1.14	2.63	4.50	−1.73	0.10	2.39
N = 300	NW ℓ_1	−0.13	0.01	0.20	−0.96	−0.55	−0.02	−0.16	0.22	0.69
	NW ℓ_2	−0.17	−0.04	0.14	−1.12	−0.71	−0.19	−1.13	−0.74	−0.25
	Shrinkage	1.42	1.87	2.43	0.22	0.85	1.63	5.08	5.97	7.08
	NL-Shrinkage	1.19	1.55	2.00	−1.18	−0.87	−0.50	0.05	0.42	0.88
	POET	2.04	2.66	3.44	0.31	1.76	3.58	0.22	2.00	4.24
N = 400	NW ℓ_1	−0.53	−0.34	−0.10	−2.26	−1.83	−1.30	−2.98	−2.56	−2.04
	NW ℓ_2	−0.52	−0.34	−0.11	−0.91	−0.48	0.05	−1.56	−1.16	−0.64
	Shrinkage	1.53	2.01	2.61	1.45	2.02	2.74	3.62	4.51	5.63
	NL-Shrinkage	1.26	1.67	2.17	0.08	0.40	0.80	2.91	3.31	3.81
	POET	1.90	2.51	3.27	−0.32	1.10	2.88	−0.47	1.19	3.28
N = 500	NW ℓ_1	0.66	0.82	1.02	−2.85	−2.51	−2.09	−1.19	−0.66	0.01
	NW ℓ_2	0.62	0.78	0.99	−1.04	−0.63	−0.11	−2.41	−1.88	−1.21
	Shrinkage	−1.11	−0.92	−0.68	3.74	4.22	4.81	4.29	4.98	5.84
	NL-Shrinkage	−1.26	−1.12	−0.95	1.62	1.94	2.34	2.53	2.86	3.27
	POET	1.13	1.53	2.03	−0.69	0.75	2.55	−1.61	0.06	2.16

Nodewise ℓ_1 and ℓ_2 procedures are abbreviated NW. The estimated constant daily fee sufficient to equate the Factor-Graphical portfolio with the competing methods. Here $\chi_{\kappa=1}$ denotes the estimated fee when investors have low risk aversion. A negative value implies that the equalizing fee is imposed on the competing method, while a positive entries imply that the fee is put on the Factor-Graphical portfolio. As such, negative values imply investor utility under the competing method exceeds that of the Factor-Graphical portfolio.

we find that the Factor-Graphical method in almost all cases provide higher utility gains than the strict Factor model. The results are much more inconclusive with regards to the Graphical and the Large-K methods. The Graphical approach in particular tend to provide high utility for this investor due to its very low turnover.

Table 4: Estimated equalization fee (χ_κ) on Factor-Graphical portfolio

		$\tilde{t} = 50$			$\tilde{t} = 150$			$\tilde{t} = 250$		
	Monthly	χ_1	χ_5	χ_{10}	χ_1	χ_5	χ_{10}	χ_1	χ_5	χ_{10}
N = 50	Large-K	−3.16	−3.03	−2.86	−2.34	−2.31	−2.26	0.49	0.58	0.69
	Factor	0.55	0.64	0.76	0.51	0.75	1.05	0.36	0.78	1.30
	Graphical	−3.18	−2.89	−2.53	−0.80	−0.37	0.17	−3.10	−2.62	−2.01
	$1/N$	−4.66	−2.23	0.83	−2.04	0.77	4.29	−2.31	0.57	4.18
N = 100	Large-K	−1.90	−1.81	−1.70	0.47	0.55	0.65	−0.42	−0.32	−0.19
	Factor	0.83	1.11	1.46	0.27	0.97	1.84	0.58	1.65	3.00
	Graphical	−3.19	−2.88	−2.48	−0.70	−0.27	0.26	−1.49	−1.00	−0.39
	$1/N$	−5.80	−3.85	−1.40	−2.07	0.34	3.38	−0.80	1.65	4.72
N = 150	Large-K	−1.75	−1.56	−1.31	−0.63	−0.61	−0.59	−1.43	−1.35	−1.24
	Factor	−0.65	−0.29	0.17	−1.21	−0.37	0.69	−0.92	0.32	1.88
	Graphical	−3.27	−2.72	−2.02	0.50	1.10	1.85	−2.62	−2.06	−1.35
	$1/N$	−4.98	−2.11	1.50	−0.15	3.11	7.18	−1.50	1.73	5.79
N = 300	Large-K	0.49	0.77	1.11	0.70	0.79	0.91	0.67	0.75	0.85
	Factor	0.55	0.91	1.37	−1.66	−0.79	0.30	1.09	2.23	3.66
	Graphical	−2.04	−1.24	−0.24	0.50	1.25	2.19	−0.47	0.27	1.21
	$1/N$	−3.30	0.41	5.05	−1.22	2.85	7.96	−0.36	3.64	8.65
N = 400	Large-K	−0.77	−0.42	0.03	−0.17	−0.03	0.15	−0.01	0.11	0.27
	Factor	−0.06	0.33	0.82	−1.97	−1.12	−0.05	1.11	2.18	3.53
	Graphical	−2.42	−1.48	−0.30	−0.76	0.12	1.23	−1.46	−0.55	0.60
	$1/N$	−3.65	0.31	5.26	−1.97	2.38	7.83	−1.36	2.93	8.30
N = 500	Large-K	−1.56	−1.30	−0.98	−0.84	−0.70	−0.52	−0.93	−0.85	−0.74
	Factor	−0.56	−0.25	0.15	−2.31	−1.51	−0.50	−0.06	0.95	2.20
	Graphical	−2.63	−1.83	−0.84	−1.45	−0.46	0.79	−3.29	−2.36	−1.20
	$1/N$	−4.26	−0.39	4.46	−2.17	2.42	8.17	−2.21	2.28	7.90

Nodewise ℓ_1 and ℓ_2 procedures are abbreviated NW. The estimated constant daily fee sufficient to equate the Factor-Graphical portfolio with the competing methods. Here $\chi_{\kappa=1}$ denotes the estimated fee when investors have low risk aversion. A negative value implies that the equalizing fee is imposed on the competing method, while a positive entries imply that the fee is put on the Factor-Graphical portfolio. As such, negative values imply investor utility under the competing method exceeds that of the Factor-Graphical portfolio.

In terms of investor utility we again find that risk averse actors still favor the Factor-Graphical method. These investors accept very high fees, between 6 and 10 percent, on the Factor-Graphical portfolio before becoming indifferent. The investors with medium levels of relative risk aversion also favour the Factor-Graphical portfolio. The low risk aversion investors prefer the equally weighted portfolio. Comparing the implied indifference fee of Table 3 to the results reported in Table 2, it is clear that the portfolio delivering the highest

utility for low risk aversion investors is the $1/N$ portfolio. For these investors, direct inverse estimation only outperforms the $1/N$ allocation under the largest available window length, in which case they prefer the nodewise ℓ_1 portfolio allocation.

5 Conclusions

In this paper we make use of statistical regularization to improve the performance of minimum variance portfolio allocation. With these methods we are able to circumvent the explicit high-dimensional matrix inversions while also reducing the impact of the noisy returns. The method we propose targets the inverse covariance matrix directly but relies on sparsity assumptions. We motivate these assumptions through a common financial factor structure. This structures partitions the minimum variance allocations into parsimonious factor exposures and sparse hedging portfolios. This is an appealing framework since it places the assumption of sparsity onto the partial correlations of residuals to the factor model. We contrast this approach under common financial structures to the alternative of using sparsity restrictions in the pairwise linear relations, which we find are more restrictive.

The method we promote adds to the large literature of methods that attempt to tackle the high-dimensional portfolio allocation, among them Ledoit and Wolf (2004), Fan et al. (2008), Fan et al. (2013) Fan et al. (2015), Goto and Xu (2015), Ledoit and Wolf (2015) and Cai et al. (2019). The crucial method we make use of is the graphical lasso of Yuan and Lin (2007) and algorithm from Friedman et al. (2008). The advantage of this approach is that it circumvents the issues associated with high-dimensional matrix inversion while remaining computationally feasible. In addition to the large-N framework we also consider the large-K, which is the large potential set of candidate factors that are available.

We evaluate the method on a data set consisting of companies that at some point since 2005 were part of the S&P500 index. Observing these assets on a daily frequency over seven years following 2010 we optimize the portfolio either every 30 days. The portfolio allocation we explore is a minimum variance strategy over three estimation windows, 50, 150 or 250 days. We compare the performance of our Factor-Graphical method against the prominent alternatives in the literature and find statistically significant volatility reductions across most portfolio dimensions. In addition we find that the gains we make are due to

the combination of financial factors and data-driven model selection, with neither tool separately able to create the same level of volatility reduction. Furthermore, we find that performing model selection among within a large set of factor candidates does not tend to further improve performance beyond what is achieved by the five Fama-French factors. This finding further emphasizes the importance of circumventing the explicit large matrix inversions when forming minimum variance portfolios.

References

Bai, J. and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

Cai, T., W. Liu, and X. Luo (2011): "A constrained 1 minimization approach to sparse precision matrix estimation," *Journal of the American Statistical Association*, 106, 594–607.

Cai, T. T., J. Hu, Y. Li, and X. Zheng (2019): "High-dimensional minimum variance portfolio estimation based on high-frequency data," *Journal of Econometrics*.

Callot, L., M. Caner, A. Özlem Önder, and E. Ulaşan (2019): "A Nodewise Regression Approach to Estimating Large Portfolios," *Journal of Business & Economic Statistics*, 0, 1–12.

Chamberlain, G. and M. Rothschild (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281–304.

Chen, N., R. Roll, and S. Ross (1986): "Economic Forces and the Stock Market," *Journal of Business*, 59, 383–403.

DeMiguel, V., L. Garlappi, and R. Uppal (2007): "Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy?" *Review of Financial Studies*, 22, 1915–1953.

Fama, E. F. and K. R. French (1993): "Common risk factors in the return of stocks and bonds," *Journal of Financial Economics*, 33, 3–56.

——— (2015): "A Five-Factor Asset-Pricing Model," *Journal of Financial Economics*, 116, 1–22.

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

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

Fan, J., Y. Liao, and X. Shi (2015): "Risks of large portfolios," *Journal of econometrics*, 186, 367–387.

Fleming, J., C. Kirby, and B. Ostdiek (2001): "The economic value of volatility

timing," *The Journal of Finance*, 56, 329–352.

——— (2003): "The economic value of volatility timing using "realized" volatility," *Journal of Financial Economics*, 67, 473–509.

Foygel, R. and M. Drton (2010): "Extended Bayesian Information Criteria for Gaussian Graphical Models," in *Advances in Neural Information Processing Systems 23*, ed. by J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, Curran Associates, Inc., 604–612.

Friedman, J., T. Hastie, and R. Tibshirani (2008): "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, 9, 432–441.

Garlappi, L., R. Uppal, and T. Wang (2007): "Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach," *Review of Financial Studies*, 20, 41–81.

Goto, S. and Y. Xu (2015): "Improving mean variance optimization through sparse hedging restrictions," *Journal of Financial and Quantitative Analysis*, 50, 1415–1441.

Hautsch, N., L. M. Kyj, and R. C. Oomen (2012): "A blocking and regularization approach to high-dimensional realized covariance estimation," *Journal of Applied Econometrics*, 27, 625–645.

Jobson, J. and B. Korkie (1980): "Estimation for Markowitz efficient portfolios," *Journal of the American Statistical Society*, 371, 544–554.

Kan, R. and G. Zhou (2007): "Optimal portfolio choice with parameter uncertainty," *Journal of Financial and Quantitative Analysis*, 42, 621–656.

Kozak, S., S. Nagel, and S. Santosh (2019): "Shrinking the cross-section," *Journal of Financial Economics*.

Ledoit, O. and M. Wolf (2003): "Improved Estimation of the Covariane Matrix of Stock Returns with an Application to Portfolio Selection," *Journal of Empirical Finance*, 10, 603–621.

——— (2004): "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of multivariate analysis*, 88, 365–411.

——— (2011): "Robust performances hypothesis testing with the variance," *Wilmott*, 2011, 86–89.

——— (2012): "Nonlinear shrinkage estimation of large-dimensional covariance matrices," *The Annals of Statistics*, 40, 1024–1060.

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

——— (2017): "Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks," *The Review of Financial Studies*, 30, 4349–4388.

Markowitz, H. (1952): "Portfolio Selection," *Journal of Finance*, 7, 77–91.

Mazumder, R. and T. Hastie (2012): "The graphical lasso: New insights and alternatives," *Electronic journal of statistics*, 6, 2125.

Meinshausen, N. and P. Bühlmann (2006): "High-dimensional graphs and variable selection with the lasso," *The annals of statistics*, 1436–1462.

Meurant, G. (1992): "A review on the inverse of symmetric tridiagonal and block tridiagonal matrices," *SIAM Journal on Matrix Analysis and Applications*, 13, 707–728.

Michaud, R. (1989): "The Markowitz optimization enigma: Is optimized optimal?" *Financial Analysts Journal*, 45, 31—42.

Ravikumar, P., M. J. Wainwright, G. Raskutti, and B. Yu (2011): "High-dimensional covariance estimation by minimizing 1-penalized log-determinant divergence," *Electronic Journal of Statistics*, 5, 935–980.

Ross, S. (1976): "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13, 341–360.

Rothman, A. J., P. J. Bickel, E. Levina, and J. Zhu (2008): "Sparse permutation invariant covariance estimation," *Electronic Journal of Statistics*, 2, 494–515.

Sharpe, W. (1964): "Capital Asset Prices: A Theory for Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19, 425–442.

Stevens, G. V. (1998): "On the inverse of the covariance matrix in portfolio analysis," *The Journal of Finance*, 53, 1821–1827.

Wainwright, M. J. (2009): "Sharp thresholds for High-Dimensional and noisy sparsity recovery using ℓ_1 -Constrained Quadratic Programming (Lasso)," *IEEE transactions on information theory*, 55, 2183–2202.

Yuan, M. and Y. Lin (2007): "Model selection and estimation in the Gaussian graphical model," *Biometrika*, 94, 19–35.

Zhao, P. and B. Yu (2006): "On model selection consistency of Lasso," *Journal of Machine learning research*, 7, 2541–2563.

A The second norm of non-equal portfolio allocations

In this section we demonstrate that for non-equal portfolio allocations the second norm is minimized by assigning two unique weights within the allocation. For an N-vector of two non-equal, non-zero, weights (w_1 and w_2) assigned to S and $(N-S)$ elements we have,

$$||w|| = \sqrt{w'w} = \sqrt{(N-S)w_1^2 + Sw_2^2}. \tag{26}$$

Forming any three weights by adjusting a subset of the elements assigned weight w_2 gives,

$$||w|| = \sqrt{w'w} = \sqrt{(N-S)w_1^2 + \tilde{S}\tilde{w}_2^2 + \hat{S}\hat{w}_2^2}, \tag{27}$$

where $S = \tilde{S} + \hat{S}$ and since the allocation should remain fully invested in risky assets we have $\tilde{S}\tilde{w}_2 + \hat{S}\hat{w}_2 = Sw_2$, which implies that $w_2 = \frac{\tilde{S}\tilde{w}_2 + \hat{S}\hat{w}_2}{S}$. Re-writing the above norm gives,

$$\begin{aligned} ||w^*|| = \sqrt{w'w} &= \sqrt{(N-S)w_1^2 + \tilde{S}\tilde{w}_2^2 + \hat{S}\hat{w}_2^2 + Sw_2^2 - Sw_2^2} \\ &= \sqrt{(N-S)w_1^2 + Sw_2^2 + (\tilde{S}\tilde{w}_2^2 + \hat{S}\hat{w}_2^2 - Sw_2^2)}. \end{aligned} \tag{28}$$

Through substitution we can write the terms within the parenthesis as,

$$\begin{aligned} \tilde{S}\tilde{w}_2^2 + \hat{S}\hat{w}_2^2 - Sw_2^2 &= \tilde{S}\tilde{w}_2^2 + \hat{S}\hat{w}_2^2 - S\Big(\frac{\tilde{S}\tilde{w}_2 + \hat{S}\hat{w}_2}{S}\Big)^2 \\ &= \frac{(\hat{S}+\tilde{S})\hat{S}\hat{w}_2^2 + (\hat{S}+\tilde{S})\tilde{S}\tilde{w}_2^2 - \hat{S}^2\hat{w}_2^2 - \tilde{S}^2\tilde{w}_2^2 - 2\hat{S}\tilde{S}\hat{w}_2\tilde{w}_2}{\tilde{S}+\hat{S}} && (29) \\ &= \frac{\tilde{S}\hat{S}\hat{w}_2^2 + \hat{S}\tilde{S}\tilde{w}_2^2 - 2\hat{S}\tilde{S}\hat{w}_2\tilde{w}_2}{\tilde{S}+\hat{S}} && (30) \\ &= \frac{\tilde{S}\hat{S}(\hat{w}_2 - \tilde{w}_2)^2}{\tilde{S}+\hat{S}} && (31) \end{aligned}$$

Since this term is greater than zero for all non-equal $\hat{w}, \tilde{w}$ we have that (26) is greater than (27). Hence, the second norm of a non-equal portfolio allocation fully invested in risky assets is minimized when there are two unique weights in the allocation.

B The graphical lasso

For the purposes of financial portfolio allocation for large N we propose solving the likelihood optimization of (15) by using the Glasso algorithm of Friedman et al. (2008). To introduce their algorithm define a partition of the inverted target covariance matrix, in this case Θ, and for later notational convenience the inverse $C = \Theta^{-1}$,

$$\Theta = \begin{pmatrix} \Theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{pmatrix}, \quad C = \begin{pmatrix} C_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}, \tag{32}$$

where Θ_{11} has dimensions $(N-1) \times (N-1)$, θ_{12} and θ_{21} are both of $(N-1) \times 1$ and θ_{22} is scalar. Analogously define the same partition of the covariance matrix and the elements-wise sign matrix, $\Gamma = \text{sign}(\Theta)$,

$$S = \begin{pmatrix} S_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix}, \quad \Gamma = \begin{pmatrix} \Gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix}. \tag{33}$$

The algorithm uses these matrix partitions to optimize penalized log-likelihood (15), $Q = -\log(\det(\Theta)) + \text{tr}(\widehat{S}\Theta) + \Lambda||\Theta||_1$. Friedman et al. (2008) start by considering the first order conditions,

$$-\Theta^{-1} + S + \Lambda\Gamma = 0. \tag{34}$$

In order to separate these conditions along the lines of the partitions of (33) it is convenient to express the partition of C as a function of the elements of Θ. Hence write this partition as,

$$C = \begin{pmatrix} (\Theta_{11} - \theta_{12}\theta_{22}^{-1}\theta_{21})^{-1} & -C_{11}\theta_{12}\theta_{22}^{-1} \\ (-C_{11}\theta_{12}\theta_{22}^{-1})' & (\theta_{22} - \theta_{21}\Theta_{11}^{-1}\theta_{12})^{-1} \end{pmatrix}. \tag{35}$$

As we separate the normal equations (34) into their respective blocks we immediately write the condition on the N:th row/column under the partition (35),

$$C_{11}\theta_{12}\theta_{22}^{-1} + s_{12} + \lambda\gamma_{12} = 0. \tag{36}$$

Here it is convenient to define $\beta = \theta_{12}\theta_{22}^{-1}$. With the normal equations rewritten in this manner Friedman et al. (2008) point out that these conditions are the optimality conditions

of the minimization problem,

$$\hat{\beta} = \text{argmin}\left(\frac{1}{2}\beta' C_{11}\beta + \beta' s_{12} + \lambda|\beta|\right). \tag{37}$$

Initializing the algorithm under some initial C they solve this optimization for an estimate $\hat{\beta}$ and update c_{12} using the partition in (35). Iteratively changing the target column the algorithm updates the estimate column-by-column until the adjustments fall below a given threshold. Upon convergence they ultimately update the estimate of the inverse, $\hat{\theta}_{22} = (c_{22} - c_{12}'\hat{\beta})^{-1}$ and $\hat{\theta}_{12} = -\hat{\beta}\hat{\theta}_{11}$. There are some known drawbacks to this algorithm.

Algorithm 1 The Glasso algrorithm (Friedman et al., 2008)

1: **Initialize:**
$C = S + \lambda\mathbf{I}$

2: **Perform following steps until convergence:**
3: Rearrange columns/rows such that the target is last.
4: Compute and save solution to (37), update row/column $c_{12} = -C_{11}\hat{\beta}$

5: For every column/row update diagonal elements θ_{22} and recover θ_{12} from $\hat{\beta}$,

$$\hat{\theta}_{22} = (c_{22} - c_{12}'\hat{\beta})^{-1}$$
$$\hat{\theta}_{12} = -\hat{\beta}\hat{\theta}_{22}$$

Mazumder and Hastie (2012) show that unless the algorithm converges there is no guarantee that the estimate of Θ will be positive semi-definite. This approach is ill-suited to situations where the we seek a computationally fast approximation of a high-dimensional inverse. This is the case where the dimensions of the data are so immense that there is no practical way of reaching convergence. It is from this observation they develop an algorithm that constructs a positive semi-definite inverse at every iteration. However, for our purposes the computational cost is not high enough to warrant an implementation of their algorithm. Therefore, we will continue under the framework outlined above.

C Illustrating the Factor-Graphical approach by simulation

To illustrate the difference between estimating the inverse of the marginal covariance matrix and the inverse of the idiosyncratic inverse we set up a simple simulation study. We use

Equation (4) to represent the returns of a set of N assets.

$$y = Bf + e, \quad f \sim \mathrm{N}(0, \Psi), \quad E \sim \mathrm{N}(0, \Delta), \quad B \sim \mathrm{U}_{(1,2)}, \tag{38}$$

where y denotes a N-vector of returns, and f is a K-vector of factors. The loading matrix, B, is $N \times K$. The covariance matrix of returns is $\Sigma = B\Psi B' + \Delta$, with elements σ_{ij}. For our purposes we set the number of factors to $K = 20$. The factor covariance matrix Ψ is a diagonal matrix with variances drawn from a uniform probability distribution from 1 to 4.5, $\mathrm{diag}(\Psi) \sim \mathrm{U}(1, 4.5)$. We select these values based on the standard deviations of the five, daily, Fama and French factors observed from from 2005 until 2015. We take the asset specific sensitivities, B, from a uniform distribution from 0 to 2. At last, we calibrate the covariance matrix of the error terms Δ. We consider a diagonal matrix where the univariate fit of the factor model provides an R^2 equal to 0.5, hence the error variances are $\delta_{ii}^2 = (\sigma_{ii}^2 - 0.5\sigma_{ii}^2)/0.5$. We draw $T = 200$ observations from this framework.

We deviate from the strict factor structure and introduce misspecification through the error covariance matrix. We let the residual correlation matrix follow symmetric Toeplitz structure. The pairwise correlation coefficients decay linearly in the log-scale to 10^{-8}. Hence, the correlation coefficients are all in the sequence $\rho_p = \exp\left(\ln(1) - \frac{(p-1)}{(N-1)} \ln\left(\frac{1}{10^{-8}}\right)\right)$ where $p = 1 \ldots N$.

$$\begin{pmatrix} 1 & \rho_2 & \rho_3 & \rho_4 & \cdots & \rho_n \\ \rho_2 & 1 & \rho_2 & \rho_3 & \cdots & \rho_{n-1} \\ \rho_3 & \rho_2 & 1 & \rho_2 & & \vdots \\ \rho_4 & \rho_3 & \rho_2 & 1 & & \vdots \\ \vdots & \vdots & & & \ddots & \vdots \\ \rho_n & \rho_{n-1} & \cdots & \cdots & \cdots & 1 \end{pmatrix} \tag{39}$$

This correlation structure implies that the inverse covariance matrix is tri-diagonal, which means that the fraction of off-diagonal elements that are non-zero tends to zero as N tend to infinity.

The important property is that the conditional correlation between assets is not all non-zero. Without that structure featuring sparse hedging portfolios is redundant beyond the common factors. Consider misspecification in the factor model. We refer to a well-specified

model as a model that includes 19 out of 20 relevant factors. The severely misspecified model instead omits 19 relevant factors.

The idiosyncratic and marginal inverses are estimated using the graphical lasso,

$$\text{argmin}_{\Theta \succeq 0} f(\Theta) = -\log(\det(\Theta)) + \text{tr}(\widehat{S}\Theta) + \Lambda||\Theta||_1, \tag{40}$$

where $\widehat{S} = T^{-1}\widehat{E}'\widehat{E}$, the covariance matrix of the residuals from (4) conditional on a set of factors. Both estimates require regularization parameters Λ. To reduce the computational cost of the simulations we calibrate the regularization using the extended Bayesian Information Criteria($eBIC$) (Foygel and Drton, 2010). Denoting the number of non-zero entries in the matrix as $\mathbf{E}$ and $\psi \in [0, 1]$ the information criteria is defined as,

$$eBIC_\psi = -2l_T\big(\Theta\big) + \mathbf{E}\log(T) + 4\psi\log(N)\mathbf{E}, \tag{41}$$

where inverse is denoted Θ and the sample covariance matrix S. The log likelihood of the inverse is $l_T\big(\Theta\big) = \frac{T}{2}\big(\log(\det(\Theta)) - \text{tr}(\Theta S)\big)$.

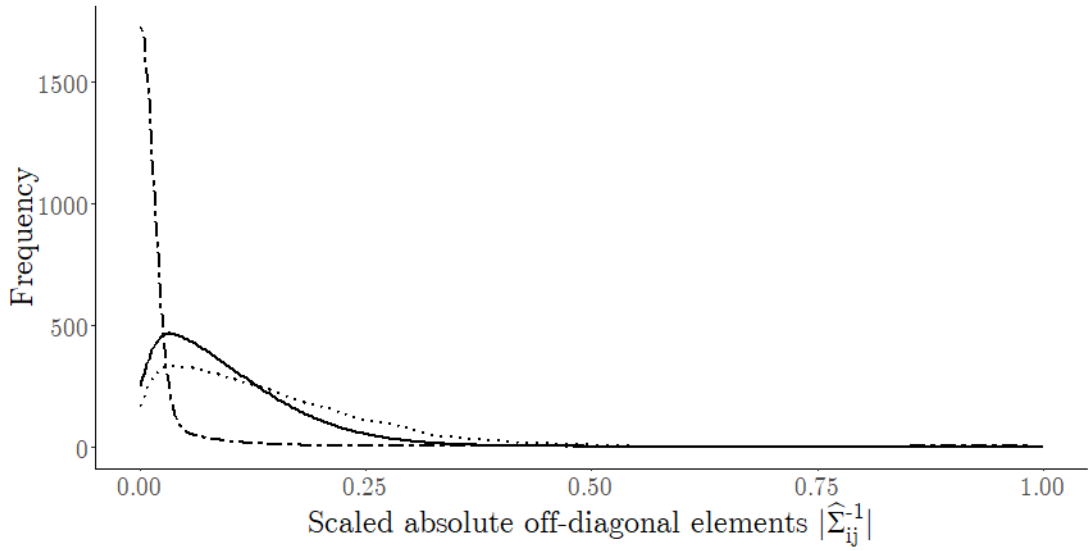
We simulate a set of $N = 100$ assets under a factor model of $K = 20$ observed factors over $T = 200$ observations.[14] The decaying correlation pattern generates a inverse covariance matrix that conditionally on the factors is sparse. Factors are drawn from independent normal distributions with the respective loadings uniformly distributed. We calibrate the error variance such that the univariate fit of the factors is equal to 50 percent. Thus, the simulation setup is not a strict factor structure but there are strong commonalities in the covariation of the series.

Using this setup it is possible to estimate the marginal inverse either directly or through the idiosyncratic inverse and (5), which is the Factor-Graphical method. Estimating the marginal inverse using both approaches we plot the distribution of off-diagonal elements and compute the accuracy of the respective estimates using the Frobenius norm. Figure 6 display the respective distributions of absolute off-diagonal elements.

The distributions are computed over $L = 1000$ simulations. The frequencies are scaled such that the counts correspond to the frequencies per simulation and the off-diagonal ele-

[14] In our empirical application we make the dimensionality issues even more severe by allowing N to far exceed T.

Figure 6: Off-diagonal elements of estimates $\hat{\Sigma}^{-1}$

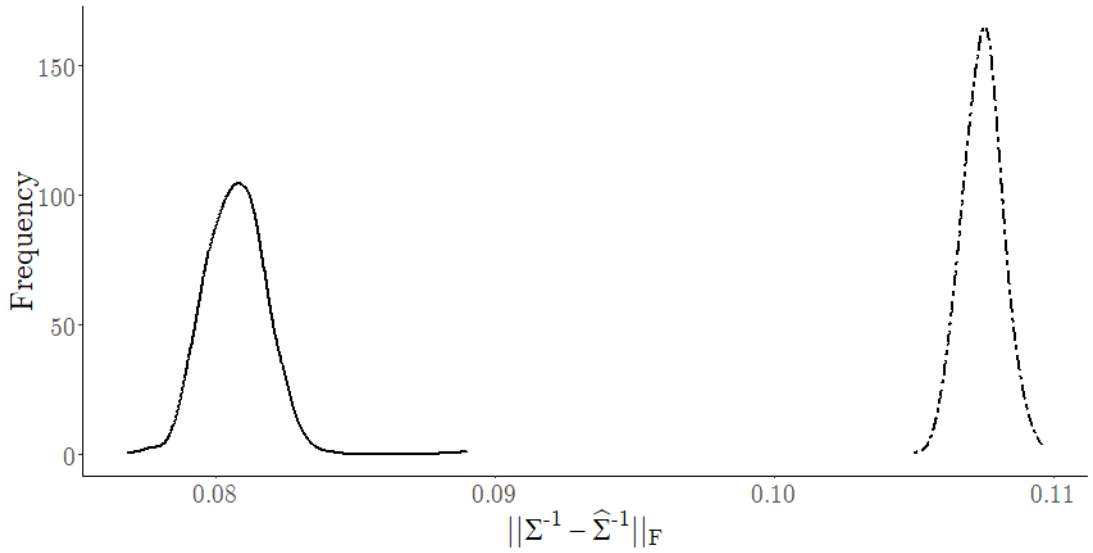


Distribution of absolute off-diagonal elements. Frequencies are averaged over $L = 1000$ simulations. Factor-Graphical (solid) and the direct marginal inverse estimate(dashed) is plotted along the actual simulated distribution (dotted). Both estimates employ regularization that promotes sparsity. Off-diagonal elements are multiplied by 1000. The simulations are performed using a fixed design.

ments are multiplied by 1000. The distribution of the off-diagonal elements in the estimate based on the idiosyncratic inverse is denoted by the solid line. This distribution follows the pattern of the actual simulated distribution pretty closely. The estimator that targets Σ^{-1} directly fares much worse. The two-dashed line denotes the off-diagonal elements of this estimator and fails to replicate the distribution of the actual data. Since this estimator regularizes Σ^{-1} using a penalization that promotes sparsity we find that the distribution of elements deviates much from the factor structure. The sparsity of the estimate is far too extreme compared to the actual distribution of off-diagonal elements. Figure 6 provides a view of the bias of the respective methods. Next we turn to the uncertainty in the estimate as measured by the standard errors. Figure 7 shows the distribution of the Frobenius distance between the estimated inverse and the actual inverse covariance matrix over the simulations.

The estimate based on the idiosyncratic inverse and the factors provides a more precise estimate of the actual inverse. It appears that the factors creates an inverse that the estimate of the direct marginal inverse cannot recover. Specifically, the common factor structure works against the sparsity assumption of the regularized estimators. The idiosyncratic

Figure 7: Fit of the estimate $\widehat{\Sigma}^{-1}$ to actual Σ^{-1}



Frobenius norm of difference between the simulated marginal inverse(Σ^{-1}) and estimates ($\widehat{\Sigma}^{-1}$). Distribution of norm over $L = 1000$ simulations using the Factor-Graphical(solid) estimate and using the direct marginal inverse estimate(dashed). Both estimates employ regularization that promotes sparsity.

inverse, however, is sparse. Therefore we find that the accuracy of the Factor-Graphical approach is always more accurate than directly estimating the marginal inverse in all of the L simulations.

In reality the factor model is unknown and to some degree misspecified, which is partly reflected in our simulation set up.[15] Clearly, if all factors of a model are redundant the Factor-Graphical approach is reduced to the method of targeting Σ^{-1} directly. However, under non-redundant factor structures the partition improves the risk reduction, increasingly so when factors fit the returns more closely.

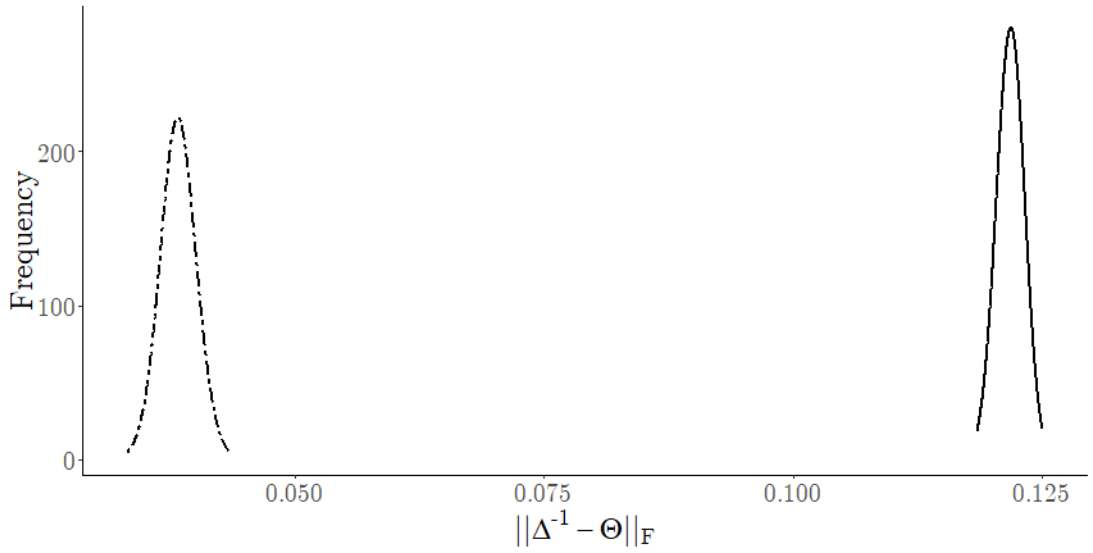
C.1 Regularization: Shrinkage to promote sparsity

While the idiosyncratic inverse appears to be a good target for regularized estimation it is not obvious which for of regularization is optimal. One type of regularization is through the form of Shrinkage, where estimates are attenuated. However, in addition to attenuation we can regularize to set estimates exactly equal to zero, in other words regularization to

[15]The issue of factor model misspecification is interesting with respect to what type of regularization is optimal. A full treatment of that issue is beyond the scope of this paper but Appendix C.1 provides a motivation for the choice we pursue here.

promote sparsity. It is not obvious which form of regularization is best suited to estimate the idiosyncratic inverse when the factor model omits relevant factors. To illustrate the issue consider the same simulation setup as before with 100 series each of 200 observations. The idiosyncratic inverse is approximated using two estimators, the first estimator induce shrinkage and sparsity, ℓ_1 regularization, while the second estimator only induce shrinkage, ℓ_2 regularization. To simplify the implementation we estimate the inverses using the neighbourhood selection approach of N nodewise regularized regressions. We consider two cases, the first case is that of a small number of omitted factors and the second case of severely misspecified models. The first model omits 1 of the 20 relevant factors. The severely misspecified model omits 19 of the factors. Repeating the simulation L times we compute the Frobenius norm of the difference between the estimated idiosyncratic inverse and the actual inverse. Figure 8 displays the norms of the two estimators, ℓ_1 and ℓ_2, using the factor model that omits one of the relevant factors. The level of regularization in the respective estimates is calibrated using 5-fold cross-validation.

Figure 8: Fit of the estimates $\widehat{\Theta}$ to Δ^{-1} with single omitted factor.

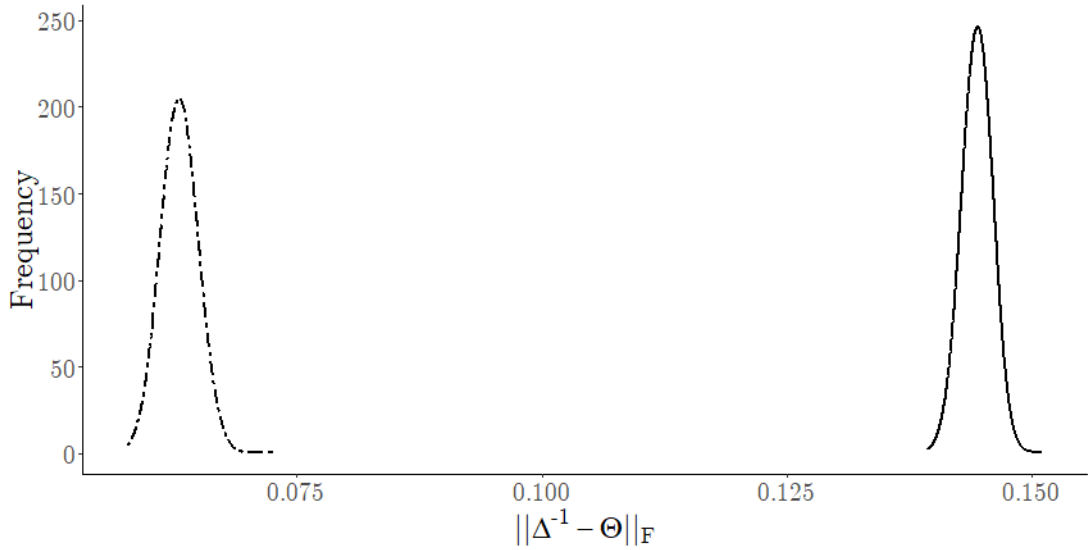


Frobenius norm of difference between the simulated idiosyncratic inverse(Δ^{-1}) and two estimates ($\widehat{\Theta}$). The estimate is either regularized with a ℓ_1 or ℓ_2 penalty. The solid line denotes the norm achieved using the ℓ_2 regularization. The two-dashed line denotes ℓ_1 regularization. Based on $L = 1000$ simulations using a fixed design.

The dashed line denotes the density of the sparse estimate while the dotted represents the shrinkage estimator. The regularization that promotes sparsity in the estimate of the

idiosyncratic inverse also achieves a better fit to the actual inverse. In relation to the ℓ_2 regularization the Frobenius norm is reduced by more than half. The second case, where the factor model omits 19 out of 20 factors is displayed in Figure 9.

Figure 9: Fit of the estimates $\widehat{\Theta}$ to Δ^{-1} with single included factor.



Frobenius norm of difference between the simulated idiosyncratic inverse and two estimates (Θ). The estimate is either regularized with a ℓ_1 or ℓ_2 penalty on the off-diagonal elements. The solid line denotes the norm achieved using the ℓ_2 regularization. The two-dashed line denotes ℓ_1 regularization. Based on $L = 1000$ simulations using a fixed design.

We find that the accuracy of the ℓ_1 regularization strictly dominates the estimator using the ℓ_2 penalization. However, the relative gain under the severally misspecified model is less compared to the first case. In all, imposing sparsity in the estimate in addition to shrinkage improves the accuracy of the estimated idiosyncratic inverse. This is found both when the factor model is severely misspecifed and when the model only omits one factor. These results guide us in the direction of the ℓ_1 penalization but we have not yet evaluated the relative portfolio performance under the two estimators. Hence, for our central empirical analysis the main focus will be on regularizations that promote sparsity. Based on these results we expect these estimators to generate the best portfolio performance, but we shall also consider estimators under ℓ_2 regularization.

D S&P500 financial data

All the data we use for the empirical exercise is collected from CRSP. We restrict the sample to assets of companies that at some point were part of the S&P500 companies between 2005 and 2017. This ensures that the stocks we allocate for our portfolios are of significant market capitalization and traded on a regular basis. Below follows basic descriptive statistics of the returns in excess to the risk-free rate including annualized volatilities and Sharpe ratios.

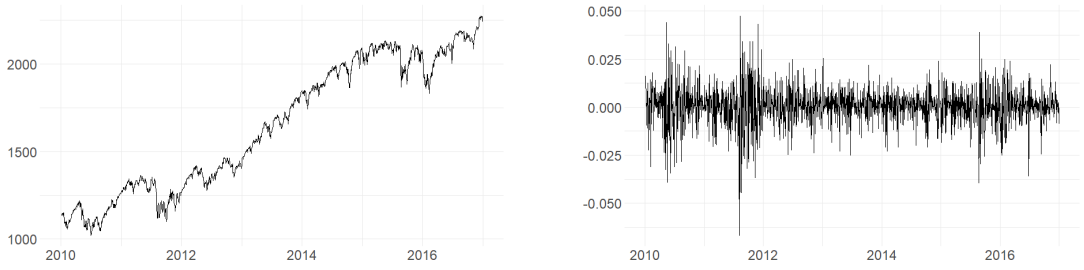
Table 5: Descriptive statistics of S&P500 firms

N = 500	Mean	St.Dev.	Min.	Q1	Q2	Q3	Max
Returns	0.16	0.08	−0.16	0.12	0.16	0.20	0.56
Volatilities	0.29	0.10	0.14	0.22	0.27	0.33	0.72
Sharpe Ratios	0.61	0.31	−0.35	0.45	0.62	0.82	1.54

Descriptive statistics of the average annualized daily returns, annualized return volatility and annualized Sharpe ratios of the S&P500 companies during 2010-2017. Mean refers to the arithmetic average, St.Dev the standard deviation, Q1, Q2 and Q3 denotes the first quartile, median and third quartile. The number of assets is denoted N.

The summary statistics displayed in Table 5 indicate that the annualized return volatilities range from 0.14 to 0.72 centered around 0.29. Table

Figure 10: S&P500 Composite Index and Return, 2010-2017



(a) Index in Level

(b) Index Return

Figure 10 shows the development of the index. It is clear that there have been significant positive development from 2012 to 2016, however, the years preceding and following that growth period saw significant volatility in the index return. Analogously to the previous descriptive statistics Table 6 below presents summary statistics of the financial anomaly returns we use as candidate predictor of the cross-section of stock returns.

Table 6: Descriptive statistics of anomaly characteristics

K = 50	Mean	St.Dev.	Min.	Q1	Q2	Q3	Max
Returns	0.01	0.06	−0.13	−0.03	−0.01	0.02	0.14
Volatilities	0.12	0.02	0.08	0.11	0.12	0.13	0.18
Sharpe Ratios	−0.08	0.44	−1.12	−0.27	−0.11	0.17	0.84

Descriptive statistics of the average annualized daily returns, annualized return volatility and annualized Sharpe ratios of the S&P500 companies during 2010-2017. Mean refers to the arithmetic average, St.Dev the standard deviation, Q1, Q2 and Q3 denotes the first quartile, median and third quartile. The number of assets is denoted N.

E Linear shrinkage, Non-linear shrinkage and POET

The shrinkage estimators of Ledoit and Wolf (2003, 2004, 2012, 2015) are among the prominent methodologies for high-dimensional covariance estimation. These estimators are based on the convex combination of the unconditional empirical covariance matrix and a prior target matrix. The target can be implied by a set of economic factors or, alternatively, a numerically well-behaved statistical construct. In this paper we make use of to shrinkage estimators adopted from Ledoit and Wolf (2004) and Ledoit and Wolf (2015). The former imposes linear shrinkage on the eigenvalues of the unconditional covariance matrix toward the mean eigenvalue. This is the same as shrinking the covariance matrix towards the identity matrix. The second method of shrinkage we implement is simlar to the former with the exception that estimated eigenvalues are non-linear transformations of the sample eigenvalues. This approach is generally found to provide improvements over its linear counterpart, both statistically and in terms of financial performance (Ledoit and Wolf, 2017). In addition to the shrinkage estimators we also consider the method of Principal Orthogonal complEment Thresholding (POET) of Fan et al. (2013). This estimator serves in our application as a statistical factor model analogue to the shrinkage estimator. The POET methodology also builds upon a parsimonious latent structure of the asset returns, but expands particularly on the case where factors are unobserved. Let $[v_1 \dots v_k]$ and $[V_1 \dots V_k]$ denote the first k eigenvalues and vectors of returns respectively. Choose the number of factors to include, $\hat{K}$, through the information criteria of Bai and Ng (2002), as is standard in the literature,

$$\hat{K} = \text{argmin}_{1 \leq K \leq K^*} \left(\ln \left(\sum_{i>K} \hat{v}_i^2 \right) + \frac{K(N+T)}{NT} \ln \left(\frac{NT}{N+T} \right) \right). \tag{42}$$

With the number of factors chosen we may express the covariance matrix as

$$S = \sum_{i=1}^{\hat{k}} v_i V_i V_i' + \sum_{i=\hat{k}+1}^{N} v_i V_i V_i' \tag{43}$$

The POET methodology replaces the latter term, $\widetilde{U} = \sum_{i=\hat{k}+1}^{N} v_i V_i V_i'$ with the term $\widehat{U}$ defined,

$$\widehat{U} = (\hat{u}_{ij})_{N\times N}, \quad \hat{u}_{ij} = \begin{cases} \tilde{u}_{ij}, & i = j \\ s_{ij}(\tilde{u}_{ij}) & i \neq j \end{cases} \tag{44}$$

One choice is to specify the function s_{ij} as a soft thresholding rule where $s_{ij} = (\tilde{u}_{ij} - \tau_{ij})_+$, where the parameter $\tau_{ij} = \zeta(\tilde{u}_{ii}\tilde{u}_{jj})^{1/2}$. The constant ζ is typically set to 0.5 when using soft thresholding rules, but can be varied over a grid until the final estimates matrix is well-conditioned. With the thresholding imposed on the residual covariance matrix we can estimate the covariance matrix of the returns.

$$\widehat{\Sigma} = \sum_{i=1}^{\hat{k}} \hat{v}_i \hat{v}_i \hat{v}_i' + \widehat{U} \tag{45}$$

This estimate is non-singular and well-conditioned. This is clearly not a comprehensive overview of the literature regarding high-dimensional covariance estimation. However, these two methodological alternatives tend to be favored in the broader portfolio allocation literature and are therefore given priority here to contrast the performance of our Factor-Graphical estimator.

F Portfolio composition

Table 7 summarizes some of the basic characteristics of the portfolios, such as *Range*, *Fraction of negative weights* and *Dispersion*. The fraction of negative weights is the sum of negative weights divided by the number of weights assigned by the respective methods. The *Range* indicates the average difference between the smallest, non-zero, weight and the largest weight in the portfolio. This overview of the distribution supply lets us discover methods prone to taking extreme positions. We measure the dispersion of the portfolio

through the sum of squared weights.

In order to be more concise we restrict our attention to the two most extreme empirical settings. The first setting being the one containing the most lenient dimensionality challenge, which is when N is 50 and the estimation window equals 250 observations. The second case is the most demanding challenge, when portfolios hold 500 assets allocated using an estimation window of 50 observations.

Table 7: Portfolio composition

	$N = 50 \quad \tilde{t} = 250$			$N = 500 \quad \tilde{t} = 50$		
	$\sum w^2$	*Range*	p_-	$\sum w^2$	*Range*	p_-
Factor-Graphical	18.96	30.32	42.63	5.40	9.58	44.01
Large-K Factor-Graphical	16.22	27.22	38.28	2.70	6.87	33.75
Nodewise ℓ_1	17.46	35.02	32.74	5.29	9.41	43.93
Nodewise ℓ_2	17.17	34.50	32.39	5.31	9.50	43.91
Shrinkage	22.16	32.43	40.37	2.60	4.92	35.43
NL-Shrinkage	19.48	30.29	38.78	2.05	4.38	33.61
POET	14.49	20.93	42.69	2.08	4.72	41.37
Factor	19.59	29.48	44.41	3.45	7.18	42.74
Graphical	5.16	11.61	18.13	1.39	3.77	32.11

The table presents summary statistics of the portfolio weights. The comparison is restricted to the empirically most demanding setting ($N = 500, \tilde{t} = 50$) and the least demanding ($N = 50, \tilde{t} = 250$). *Range* denotes the average difference between the smallest non-zero weight and the largest. The average proportion(%) of negative weights assigned by the estimator is denoted p_- and the mean sum of squared weights is $\sum w^2$. The columns $\sum w^2$ and *Range* are multiplied by 100.

The portfolios exhibit large similar characteristics across methods with the exception of the Graphical method. This estimation imposes much more strict sparsity in the estimated inverse covariance matrix and thus creates much less dispersion in the portfolio weights. The fraction of negative positions in this portfolio is also much lower than the alternatives. Apart from this special case, portfolio exhibit largely comparable dispersion, range and negative weights. The number of short positions varies between 32 and 42 percent.

Through Table 7 we also find that the none of the methods appear more or less likely to produce negative weights. The fraction of negative weights is, for all methods, around 35 percent and largely stable across sample sizes and with only marginal differences between estimators.

G Conditional sparsity in the inverse

We deviate from previous methods that emphasize conditional sparsity in the error covariance matrix by shifting the assumption to a conditionally sparse inverse covariance matrix. In this section we outline an example of a model under which our assumption provides a better representation than does the traditional sparse covariance assumption. We expand model (4) to include a set of k unobserved local factors f that provides smaller blocks of co-varying returns,

$$y = \alpha + BF + bf + u. \tag{46}$$

The covariance matrix, Δ, of the error term is diagonal. For simplicity assume that one asset only is exposed to one local factor. Thus, columns of the exposures b contain many zeroes and multiple factors do not load onto the same asset. Those returns exposed to a local factor have a loading of one. Thus, b contains only zeroes and ones. The k factors thus load onto disjoint subsets of the assets. Thus, we refer to them as local factors. We define the covariance matrix of the local factors as decaying in the off-diagonal elements such that,

$$\Psi_f = \begin{pmatrix} 1 & c_{1,2} & \dots & c_{1,k} \\ c_{2,1} & 1 & \dots & c_{2,k-1} \\ \vdots & \vdots & \ddots & \vdots \\ c_{2,k} & c_{k-1,2} & \dots & 1 \end{pmatrix}. \tag{47}$$

Where $c_{i,j} = c^{|i-j|}$ where $i, j = 1 \dots k$ and $0 < c < 1$. This is known as a Kac-Murdock-Szegö matrix. Hence, assets are exposed to local factors where the factors are correlated but in a decreasing fashion. While the covariance decay along the off-diagonal elements, the inverse Ψ_f^{-1} is tridiagonal. With this structure, for a small number of local factors, the covariance matrix $\Sigma_e = b'\Psi_f b + \Delta$ is only approximately sparse as c approach zero. The inverse covariance matrix,

$$\Delta^{-1} = (b'\Psi_f b + \Delta)^{-1} = \Delta^{-1} - \Delta^{-1} b(b'\Delta^{-1} b + \Psi_f^{-1})^{-1} b'\Delta^{-1} \tag{48}$$

The local factors create block structures within the financial assets. Since the correlation between the local factors is decreasing so is also the correlation between the blocks of assets. Ultimately, both Σ_e and Σ_e^{-1} are full matrices, however, the rate of decrease across the off-diagonal elements of the two matrices are potentially very different. The decay of the elements away from the diagonal is determined by the decay in the term $(b'\Delta^{-1}b+\Psi_f^{-1})^{-1}$,

$$(b'\Delta^{-1}b+\Psi_f^{-1})^{-1}=\begin{pmatrix}\sum b_1^2+\frac{1}{1-c^2} & \frac{-c}{1-c^2} & & \\ \frac{-c}{1-c^2} & \sum b_2^2+\frac{1+c^2}{1-c^2} & \ddots & \\ & \ddots & \ddots & \frac{-c}{1-c^2} \\ & & \frac{-c}{1-c^2} & \sum b_N^2+\frac{1}{1-c^2}\end{pmatrix}^{-1}. \tag{49}$$

The decrease across off-diagonal elements in the covariance is constant at c^{-1}, when returns belong to different local factors. In Σ_e^{-1} the rate of decrease is non-constant. Following Meurant (1992) the non-constant rates of decay in this matrix is,

$$\frac{\left((b'\Delta^{-1}b+\Psi_f^{-1})^{-1}\right)_{i,j}}{\left((b'\Delta^{-1}b+\Psi_f^{-1})^{-1}\right)_{i,j+1}}=\frac{(1-c^2)d_{j+1}}{c}. \tag{50}$$

The term $d_{\cdot}$ is the diagonal elements of a upper triangular decomposition, which here is expressed as a recursion along with a terminating condition,

$$d_i=\begin{cases}\sum b_i^2+\frac{1+c^2}{1-c^2}-\frac{(c/(1-c^2))^2}{d_{i+1}} & i<N \\ \sum b_i^2+\frac{1}{1-c^2} & i=N\end{cases}. \tag{51}$$

Following the results of Meurant (1992) we find that the elements of our inverse is strictly decreasing. Together the rate in (50) along with the term (51) gives us the rate of decrease in Δ^{-1},

$$rate_{j+1}=\frac{\left((b'\Delta^{-1}b+\Psi_f^{-1})^{-1}\right)_{i,j}}{\left((b'\Delta^{-1}b+\Psi_f^{-1})^{-1}\right)_{i,j+1}}=c^{-1}\Big(1+(1-c^2)\sum b_{j+1}^2+c^2-\frac{c^2/(1-c^2)}{d_{j+1}}\Big). \tag{52}$$

From this rate we find that the off-diagonal decay in the inverse covariance matrix always exceeds the decrease in the covariance matrix if $d_{j+1}>1/(1-c^2)$. This clearly holds for

$j = N - 1$. With the rate of decrease in the covariance matrix of c^{-1} we find,

$$rate_{j+1} > c^{-1} \iff d_{j+1} > \frac{1}{1-c^2} \tag{53}$$

Starting from the end of the recursion we take $j = N - 1$,

$$d_{N-1} = \sum b_{N-1}^2 + \frac{1+c^2}{1-c^2} - \frac{(c/(1-c^2))^2}{d_N}. \tag{54}$$

Applying the necessary inequality we find,

$$\sum b_{N-1}^2 + \frac{1+c^2}{1-c^2} - \frac{\left(c^2/(1-c^2)\right)^2}{d_N} > \frac{1}{1-c^2}. \tag{55}$$

We can simplify the formulation to,

$$(1-c^2)\sum b_{N-1}^2 + 1 + c^2 - \frac{c^2/(1-c^2)}{d_N} > 1, \tag{56}$$

.

The inequality holds if,

$$c^2 - \frac{c^2/(1-c^2)}{d_N} > 1, \tag{57}$$

$$d_N > \frac{1}{(1-c^2)}. \tag{58}$$

Since $d_N > 1/(1-c^2)$ for all non-zero local exposures we have that $d_{N-1} > 1/(1-c^2)$. Along the same lines for $j = N - 2$,

$$d_{N-2} = \sum b_{N-2}^2 + \frac{1+c^2}{1-c^2} - \frac{(c/(1-c^2))^2}{d_{N-1}} \tag{59}$$

and,

$$(1-c^2)\sum b_{N-2}^2 + 1 + c^2 - \frac{c^2/(1-c^2)}{d_{N-1}} > 1 \tag{60}$$

we require,

$$d_{N-1} > \frac{1}{(1-c^2)}. \tag{61}$$

Which we know holds from the previous case. Hence for all $j = 1, \ldots, N$ the rate of decrease in the inverse covariance matrix exceed c^{-1}, the rate in the covariance matrix.

Reviewing the off-diagonal decrease as a function of the parameter c^{-1} we find that as c tends to zero the rates in the two matrices is identical. However, as c approaches one, the difference between the rates increase in favour of the inverse. Similarly difference between the rates of decrease also grows as the exposure to the local factor increase increase.

This implies that while both the covariance and inverse covariance are full matrices, the elements of the inverse tend to zero faster as N increase. Thus, the local factors of this form provides a circumstance where the sparsity assumption is more suitable in the inverse rather than the covariance matrix. Hence, a sparse estimate of the inverse covariance matrix is more suitable than a sparse representation of the covariance matrix.