# Sparse Portfolio Selection via the sorted $\ell_1$ - Norm

Philipp J. Kremer*[1], Sangkyun Lee†[2], Małgorzata Bogdan‡[3], and Sandra Paterlini§[1]

[1]EBS Universität für Wirtschaft und Recht, Chair of Financial Econometrics and Asset Management, Gustav-Stresemann-Ring 3, 65189 Wiesbaden
[2]Division of Computer Science, College of Computing, Hanyang University ERICA, Ansan, Republic of Korea
[3]Department of Mathematics, University of Wroclaw, Wrocław, Poland

January 2, 2018

**Abstract**

We introduce a financial portfolio optimization framework that allows us to automatically select the relevant assets and estimate their weights by relying on a sorted $\ell_1$-Norm penalization, henceforth SLOPE. Our approach is able to group constituents with similar correlation properties, and with the same underlying risk factor exposures. We show that by varying the intensity of the penalty, SLOPE can span the entire set of optimal portfolios on the risk-diversification frontier, from minimum variance to the equally weighted. To solve the optimization problem, we develop a new efficient algorithm, based on the Alternating Direction Method of Multipliers. Our empirical analysis shows that SLOPE yields optimal portfolios with good out-of-sample risk and return performance properties, by reducing the overall turnover through more stable asset weight estimates. Moreover, using the automatic grouping property of SLOPE, new portfolio strategies, such as SLOPE-MV, can be developed to exploit the data-driven detected similarities across assets.

*Keywords:* Portfolio Management, Markowitz Model, Sorted $\ell_1$-Norm Regularization; Alternating Direction Method of Multipliers

*EFM Classification Codes:* 330, 370

*philipp.kremer@ebs.edu (Presenting Author)
†sangkyun@hanyang.ac.kr
‡Malgorzata.Bogdan@pwr.edu.pl
§sandra.paterlini@ebs.edu

# 1 Introduction

The development of successful asset allocation strategies requires the construction of portfolios that perform well out-of-sample, provide diversification benefits, and are cheap to maintain and monitor. When setting up quantitative portfolio selection strategies, the problem is then one of statistical model selection and estimation, that is the identification of the assets in which to invest and the determination of the optimal weight for each asset. In 1952, Harry Markowitz laid the foundation for modern portfolio theory by introducing the mean-variance optimization framework. Assuming that asset returns are normally distributed, such model requires only two input estimates: the vector of expected returns and the expected covariance matrix of the assets. Solving the quadratic optimization problem, by minimizing the portfolio expected risk, for a given level of expected return, the investor can then find the optimal portfolio allocation. Although Markowitz's model has been widely criticized, it is the backbone of the vast majority of portfolio optimization frameworks and is still largely used in practice, especially in fintech companies as part of their robo-advisory (see e.g. Kolm et al. (2014)). One of the major shortcomings of the mean-variance approach is the fact that optimized weights are highly sensitive to estimation errors and to the presence of multicollinearity in the inputs. In particular, it is acknowledged that estimating the expected returns is more challenging, than just focusing on risk minimization and thereby looking for the portfolios with minimum risk, i.e. the so-called global minimum variance portfolios (GMV) (Merton 1980, Chopra and Ziemba 1993, Jagannathan and Ma 2003). But even in the GMV set-up, the sample covariance matrix might exhibit estimation error that can easily accumulate, especially when dealing with a large number of assets (Michaud 1989, Ledoit and Wolff 2003, DeMiguel and Nogales 2009, Fan et al. 2012).

Furthermore, as a result of multicollinearity and extreme observations, the Markowitz set-up often leads to undesirable and unrealistic extreme long and short positions, which can hardly be implemented in practice, due to regulatory and short selling constraints (Shefrin and Statman 2000, DeMiguel et al. 2009b, Boyle et al. 2012, Roncalli 2013). An ideal portfolio then has conservative asset weights, which are stable in time, to avoid high turnover and transaction costs, while still promoting the right amount of diversification and being able to control the total amount of shorting.

A natural approach to solve this problem is to extend the Markowitz optimization framework by using a penalty function on the weight vector, typically given by the norm, and whose intensity is controlled by a tuning parameter $\lambda$. Probably, the most recent successful approach using convex penalty functions, that can explicitly control for the total amount of shorting, while avoiding to invest in the entire asset universe, is the Least Absolute Shrinkage and Selection Operator (LASSO) introduced by Tibshirani (1996).

In the portfolio context, the LASSO framework typically relies on adding to the Markowitz formulation a penalty proportional to the $\ell_1$-Norm[1] on the asset weight vector (Brodie et al. 2009, DeMiguel et al. 2009a, Carrasco and Noumon 2012, Fan et al. 2012). DeMiguel et al. (2009a) provide a general framework that nests regularized portfolio strategies based on the $\ell_1$-Norm with the approaches introduced by Ledoit and Wolff (2003) and Jagannathan and Ma (2003). Furthermore, the authors advocate their superior performance in an out-of-sample setting. Brodie et al. (2009) and Fan et al. (2012) show that the LASSO (a) results in constraining the gross exposures, (b) can be used to implicitly account for transaction costs, and (c) sets an upper bound on the portfolio risk depending just on the maximum estimation error of the covariance matrix. Moreover,

[1]Let $\boldsymbol{w} = [w_1, w_2, ...., w_k]'$ be the portfolio weight vector, then the $\ell_q$-Norm is defined as: $\|\boldsymbol{w}\|_q = (\sum_{i=1}^{k} |w_i|^q)^{\frac{1}{q}}$, with $0 < q < \infty$. If $q = 1$, then $\ell_1 = \sum_{i=1}^{k} |w_i|$ (LASSO), while for $q = 2$ we have $\|\boldsymbol{w}\|_2 = (\sum_{i=1}^{k} w_i^2)^{1/2}$ (RIDGE). Note that $\ell_q$ with $0 < q < 1$ is not a norm but a quasi norm.

the shrinkage covariance estimation of Jagannathan and Ma (2003), obtained by adding a no-short sale constraint (the so-called GMV long-only (GMV-LO)), can be considered a special case of the LASSO.

Next to the LASSO penalty, Brodie et al. (2009) and DeMiguel et al. (2009b) also consider a portfolio with an $\ell_2$-Norm penalty on the weight vector, known in statistical literature as RIDGE (Hoerl and Kennard 1988). Although, the RIDGE penalty stabilizes the mean-variance optimization, as it controls for multicollinearity, the shape of the penalty does not promote sparsity, leading to portfolios with an undesirably large number of active positions (Carrasco and Noumon 2012).

Despite its appealing properties, the LASSO has reported shortcomings of (a) large biased coefficient values (Gasso et al. 2010, Fastrich et al. 2015), of (b) reduced recovery of sparse signals when applied to highly dependent data, like crisis periods (Giuzio and Paterlini 2016), and of (c) randomly selecting among equally correlated coefficients (Bondell and Reich 2008). Moreover, it is ineffective in the presence of no short selling (i.e. $w_i \geq 0$) and an imposed budget constraint (i.e., $\sum_{i=1}^{k} w_i = 1$), as the $\ell_1$-Norm is then just equal to 1.

To overcome these limitations, non-convex penalties, like the $\ell_q$-Norm, the LOG and the SCAD penalty have recently gained increased attention in the portfolio literature (Gasso et al. 2010, Fastrich et al. 2014, 2015, Xing et al. 2014, Chen et al. 2016). Among these penalties, the $\ell_q$-Norm with $0 < q < 1$ has shown remarkable theoretical and empirical properties, compared to the classic LASSO (Saab et al. 2008, Fastrich et al. 2014, Chen et al. 2016, Giuzio and Paterlini 2016). In particular, Fernholtz et al. (1998) show that the $\ell_q$-Norm can be interpreted as a portfolios measure of diversification. As such, the $\ell_q$ - Norm attains its maximum for an equally weighted portfolio and its minimum for a portfolio completely invested in one single asset. The resulting allocations outperform

those based on the LASSO, especially in the presence of highly dependent data. However, from an optimization standpoint, adding non-convex penalties to the minimum variance framework results in optimization problems that are NP-hard. Thus, solutions are typically computed by relying on heuristics or local optimizer, which might not be efficient and are not guaranteed to converge to the global optimum.

In this paper, we focus on the class of convex penalty functions and extend the literature on regularization methods in various ways:

First, we introduce the *Sorted* $\ell_1$ *Penalized Estimator* (SLOPE), as a new penalty function within the mean-variance portfolio optimization framework. The SLOPE penalty takes the form of a sorted $\ell_1$ - Norm, in which each asset weight is penalized individually using a vector of tuning parameters, $\boldsymbol{\lambda}_{SLOPE} = (\lambda_1, \lambda_2, \ldots, \lambda_k)$, where $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_k \geq 0$. This sequence of $\boldsymbol{\lambda}_{SLOPE}$ values is decreasing, and the largest weight corresponds to the highest regularization parameter, such that SLOPE penalizes the weights according to their rank magnitude. In a 2-dimensional setting the penalty takes an octagonal shape (see Figure 1), is singular at the origin and promotes the grouping of variables, that is some asset weights are assigned the same coefficient value. The penalty thus combines the two favorable properties of the $\ell_1$-Norm and the $\ell_\infty$-Norm[2], which promote sparsity and the grouping of variables, respectively. So far, the study of SLOPE has mostly focused on orthogonal settings and on genetic applications, where it is used to choose relevant genes from a large group of possible explanatory factors. Our work shows that in portfolio optimization, together with an added budget constraint (i.e. $\sum_{i=1}^{k} w_i = 1$), SLOPE continues to shrink the active weights, even when short sales are restricted (i.e. $w_i \geq 0, \ \forall \ i = 1, ..., k$). Consequently, it spans the diversification frontier from the GMV-LO up to the equally weighted (EW) portfolio.

[2]Given a weight vector $\boldsymbol{w}$ with $k$ elements, the $\ell_\infty = \|\boldsymbol{w}\|_\infty = max(w_1, ....., w_k)$.

Second, we introduce a new optimization algorithm to solve the mean-variance portfolio problem with the sorted $\ell_1$ regularization and linear constraints on the asset weights. The algorithm uses the ideas of variable splitting and the Alternating Direction Method of Multipliers (ADMM) framework (Powell 1969, Hestenes 1969, Boyd et al. 2011). Using a mathematically equivalent reformulation of the original problem, the algorithm can use existing implementations of proximal operators (Parikh and Boyd 2014), associated with the $\ell_1$, the sorted $\ell_1$, and even other regularizers. Furthermore, Section 6.1 of the Appendix shows that the ADMM provides a more efficient alternative for solving the LASSO optimization problem, then the state-of-art Cyclic Coordinate Descend (CyCoDe) algorithm.

Third, we are, to our knowledge, the first to investigate the properties of SLOPE under a realistic factor model, which assumes that all assets can be represented as linear combination of a small number of hidden risk factors, as e.g. in Fan et al. (2008). In the set-up of classical multiple regression, in which the explanatory variables are assumed independent, Bogdan et al. (2013, 2015) and Su and Candès (2016) provide extensive evidence of SLOPE's superior model selection and estimation properties. Further evidence for these properties are provided by the results of Bellec et al. (2016a) and Bellec et al. (2016b), which show that contrary to LASSO, SLOPE is asymptotically optimal for the general class of design matrices satisfying the modified Restricted Eigenvalue condition.

Moreover, Bondell and Reich (2008) and Figueiredo and Nowak (2014) investigate the properties of SLOPE and its predecessor OSCAR (Octagonal Shrinkage and Selection Operator, Bondell and Reich (2008)) in the situation, when regressors are strongly correlated. Bondell and Reich (2008) apply OSCAR to agricultural data, showing that the method successfully forms predictive clusters, which can then be analyzed according to

their individual characteristics. Figueiredo and Nowak (2014) illustrate the "clustering" properties of the *ordered weighted* $\ell_1$ *- Norm* (OWL) in the linear regression framework with strongly correlated predictors, providing further simulation and theoretical results. However, none of these works addresses the interesting situation, in which the correlation structure results from the dependency of the explanatory variables on a few hidden factors and on financial real-world data.

Recently, Xing et al. (2014) applied the OSCAR to the mean-variance portfolio optimization, together with a linear combination of the $\ell_1$- and the $\ell_\infty$-Norms. They advocate the method for its ability to identify portfolios that attain higher Sharpe Ratios and lower turnovers compared to those resulting from traditional approaches like the GMV and the GMV-LO portfolios. However, they do not point out the clumping property of the OSCAR. With SLOPE, we consider a generalized framework that nests the GMV, the GMV-LO, the LASSO, the $\ell_\infty$ - Norm and the approach of Xing et al. (2014).

In this paper, we analyze the properties of SLOPE, with both simulated and real world data. The purpose of the simulations is to investigate the properties of our new penalty, when the data generating mechanism is completely known, so that the results can be compared to the so-called *oracle* solution. These simulations show that SLOPE reduces the estimation errors in the portfolio weights and groups assets depending on the same risk factors together. This grouping behavior then allows the investor to select individual constituents from the clusters, for example based on her preferences and asset-specific properties, enabling her to develop new investment strategies such as SLOPE-MV, which we introduce in Section 4.1.

For the real world data analysis, we use monthly returns of the 10- and 30-Industry portfolios (Ind), as well as the 100 Fama French (FF) portfolios formed on Size and Book-to-Market, and which cover the period from 1970 to 2017. Furthermore, we con-

sider daily returns of the S&P 100 (SP100) and S&P 500 (SP500) from 2004 to 2016. Our results show that the risk of the SLOPE portfolio is comparable to or smaller than the risk of the LASSO portfolio. Also, we observe that SLOPE outperforms the LASSO, yielding better risk- and weight diversification measures. In fact, the sorted $\ell_1$- Norm is able to span the entire risk-diversification frontier, starting from the GMV, via the GMV-LO up to the EW. The investor can then select the portfolio with the risk-diversification trade-off that best fits her preferences.

The above mentioned characteristics establish SLOPE as a new attractive portfolio construction alternative, capable of controlling short sales and identifying groups of assets. It thereby offers the possibility to implement individual views, which goes beyond the standard statistical shrinkage or regularization approaches.

The paper is structured as follows: Section 2 introduces our methodology and discusses the properties of SLOPE. Section 3 analyses the behavior of SLOPE in simulated environments, while Section 4 focuses on the empirical results. Section 5 concludes.

# 2 Sparse Portfolio Selection via the Sorted $\ell_1$-Norm

Markowitz (1952) pioneered the idea that investors should consider both risk and return, instead of just focusing on those assets that offered the largest increase in value given today's prices. Central to his argument is the notion of diversification, i.e. not only the individual securities' risk are important, but also their relationship with other assets that is how does the performance of an asset moves with or against the performance of the other assets in the market.

Then, given $k$ jointly normally distributed asset returns $R_1, \ldots, R_k$, with expected value vector $\boldsymbol{\mu} = [\mu_1, ..., \mu_k]'$ and covariance matrix $\boldsymbol{\Sigma}$, the Markowitz (1952) portfolio selec-

tion problem can be stated as the following optimization:

$$\begin{aligned} \min_{\boldsymbol{w}\in\mathbb{R}^k} \ & \frac{\phi}{2}\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} - \boldsymbol{\mu}'\boldsymbol{w} \\ \text{s.t.} \ & \sum_{i=1}^{k} w_i = 1 \end{aligned} \tag{1}$$

where $\sigma_p^2 = \boldsymbol{w}\boldsymbol{\Sigma}\boldsymbol{w}$ is the portfolio risk, $\boldsymbol{\mu}'\boldsymbol{w}$ is the portfolio return and $\phi > 0$ is the coefficient of relative risk aversion (Markowitz 1952, Fan et al. 2012, Li 2015).

Despite the advantage of being a quadratic optimization problem, the standard Markowitz model is often criticized, as it leads to extreme and unstable optimal portfolio weights. This results, on the one hand, from the multicollinearity of returns, which increases especially during crisis periods, and, on the other hand, from input changes, due to extreme data, leading to the consequent accumulation of estimation error in the weights. Besides shrinkage and robust estimation methods (see e.g., Ledoit and Wolff (2004), Welsch and Zhou (2007), Kolm et al. (2014)), one of the most recent and interesting approaches is based on statistical regularization. This approach modifies the optimization problem (1), by adding a penalty function, $\rho_\lambda(\boldsymbol{w})$, that depends on the norm of the asset weights vector, and which is typically chosen to promote sparsity in the optimal weight vector. An additional parameter $\lambda$ then controls the impact of $\rho(\boldsymbol{w})$ and thereby the amount of shrinkage applied to the weights vector and the level of sparsity.

The optimization problem can be stated as:

$$\begin{aligned} \min_{\boldsymbol{w}\in\mathbb{R}^k} \ & \frac{\phi}{2}\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} - \boldsymbol{\mu}'\boldsymbol{w} + \rho_\lambda(\boldsymbol{w}) \\ \text{s.t.} \ & \sum_{i=1}^{k} w_i = 1 \end{aligned} \tag{2}$$

The simplest approach is the LASSO, which considers as a penalty function the $\ell_1$-Norm of the asset weights vector ($\rho_\lambda(\boldsymbol{w}) = \lambda \times \sum_{i=1}^{k} |w_i|$, with $\lambda$ being a scalar). The resulting optimization problem is still convex, while promoting model selection and estimation in a single step. From a financial perspective, LASSO is interpreted as a gross exposure constraint (i.e. a constraint on the total amount of shorting) or a way to account for transaction costs (Brodie et al. 2009). However, it is not effective in the presence of both a budget ($\sum_{i=1}^{k} w_i = 1$), and a no-short selling (i.e., $w_i \geq 0$) constraint, as the $\ell_1$-norm is then simply equal to 1.

Following, we propose a more general approach that within a single optimization algorithm allows us to encompass the original LASSO, the OSCAR of Bondell and Reich (2008), and the combination of $\ell_1$ and $\ell_\infty$ penalties, as proposed in Xing et al. (2014).

In fact, we penalize the weights vector by considering as $\rho_\lambda(\boldsymbol{w})$ the sorted $\ell_1$-Norm, defined as:

$$\begin{aligned} \rho_\lambda(\boldsymbol{w}) &:= \sum_{i=1}^{k} \lambda_i |w|_{(i)} = \lambda_1 |w|_{(1)} + \lambda_2 |w|_{(2)} + ... + \lambda_k |w|_{(k)} \\ \text{s.t. } & \lambda_1 \geq \lambda_2 \geq \ldots \lambda_k \geq 0 \text{ and } |w|_{(1)} \geq |w|_{(2)} \geq \ldots |w|_{(k)} \ , \end{aligned} \tag{3}$$

where $|\boldsymbol{w}|_{(i)}$ denotes the $i$th largest element in absolute value of the vector $\boldsymbol{w}$. The sorted

$\ell_1$-Norm was originally introduced in Bogdan et al. (2013, 2015) to construct the Sorted $\ell_1$ Penalized Estimator for the selection of explanatory variables in the multiple regression model. It was also developed independently by Zeng and Figueiredo (2014) as Ordered Weighted $\ell_1$ Norm (OWL). To our knowledge, this is the first work in financial portfolio selection that applies SLOPE and discusses its grouping properties, while also introducing a new optimization algorithm.

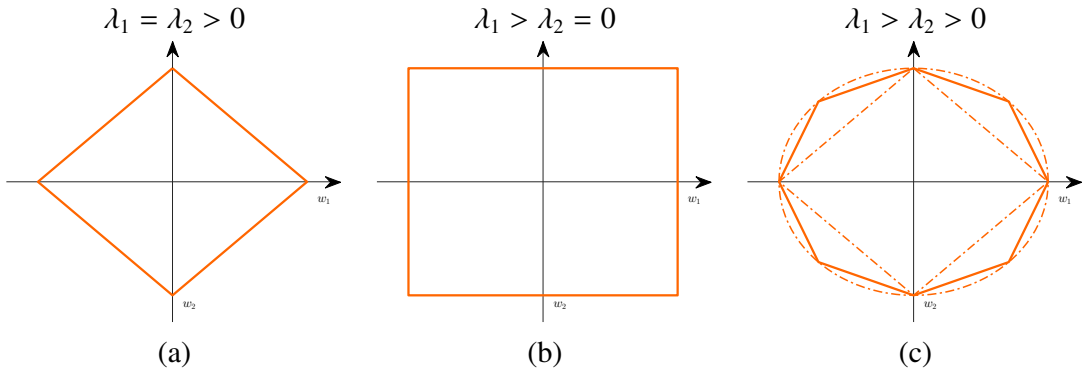
## 2.1 Geometric Interpretation

Compared to most of the other regularization methods, SLOPE does not rely on a single tuning parameter $\lambda$, but rather on a non-increasing sequence $\boldsymbol{\lambda}_{SLOPE} = (\lambda_1, \lambda_2, \ldots, \lambda_k)$, with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k \geq 0$. This sequence is aligned to the sorted weight vector, such that the largest absolute weight is penalized with the largest tuning parameter. Consequently, the sequence of $\lambda$ parameters gives a natural interpretation of importance to the asset weights, besides providing full flexibility in recapturing the profiles of the $\ell_1$- and $\ell_\infty$- Norms, as well as of their linear combinations. Figure 1 shows a simple set-up with two assets and the respective shapes of spheres (i.e. the set points for which $\rho_\lambda(\boldsymbol{w}) = c$) that we obtain, depending on how the sequence $\boldsymbol{\lambda}_{SLOPE} = (\lambda_1, \lambda_2)$ is chosen. As shown in Panel (a), if all tuning parameters have the same value, while still being larger than zero (i.e. $\lambda_1 = \lambda_2 > 0$), the SLOPE sphere coincides with the well studied diamond shape of the LASSO penalty. Through its singularity at the origin, the LASSO promotes sparse solutions that set one of the two assets' weights exactly equal to zero. On the other hand, choosing $\lambda_2 = 0$ and $\lambda_1 > 0$, yields the regularization term of the $\ell_\infty$-Norm. The respective shape, as shown in Panel (b), takes the form of a square and promotes the grouping of variables, i.e. it encourages solutions under which both asset

weights are assigned exactly the same value.

Given these two extreme cases, Panel (c) of Figure 1 shows the octagonal shape of SLOPE, obtained by using a decreasing sequence of lambda values, with $\lambda_1 > \lambda_2 > 0$. The penalty combines both properties of the LASSO and the $\ell_\infty$ penalties and due to its singularity, is either able to set some weights exactly equal to zero, and/or to assign the same value to some of the other weights. Furthermore, it approximates the already well studied RIDGE penalty, which corresponds to a circle in the 2-dimensional set-up. Although RIDGE is still convex, the shape of the penalty does not promote sparsity among the coefficients, leading to undesirable portfolios with a large number of active positions (Carrasco and Noumon 2012, DeMiguel et al. 2009a). Thus, the choice of the lambda sequence for SLOPE provides the investor with the flexibility to choose any of these shapes of the unit sphere and of the corresponding mode of shrinking the dimension of the weight vector.

Figure 1: Geometric Representation of Penalty Functions




For two asset weights $\boldsymbol{w} = [w_1\ w_2]'$, the figure shows the unit spheres for different SLOPE sequences: (a) the LASSO $\ell_1$ sphere, when $\lambda_1 = \lambda_2 > 0$, (b) the $\ell_\infty$ sphere, when $\lambda_1 > \lambda_2 = 0$ and (c) the SLOPE sphere, when $\lambda_1 > \lambda_2 > 0$. The dashed lines in (c) represent the diamond shape of the LASSO and the RIDGE $\ell_2$-balls, respectively.

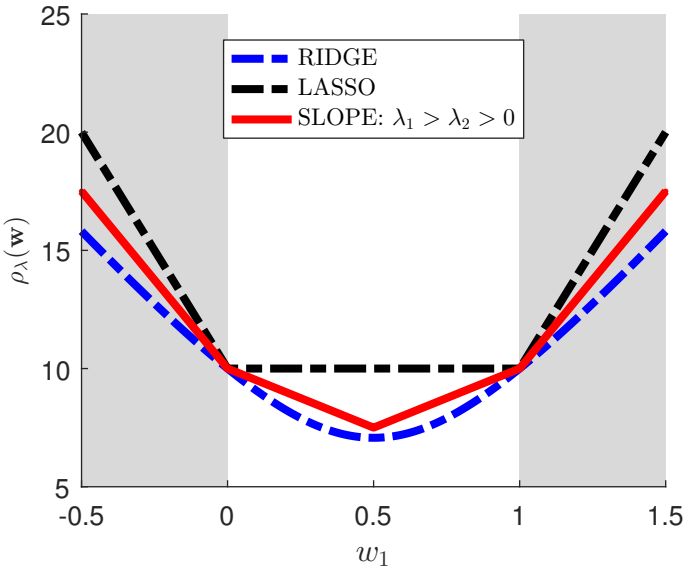
In portfolio optimization, a budget constraint that requires the weights of the portfolio

to sum to one, is imposed. Consequently, we need to understand how the penalties behave in the presence of such an additional constraint. Figure 2 plots the SLOPE penalty, together with the LASSO and the RIDGE penalty for a universe of two assets and under the condition that $w_1 + w_2 = 1$. Furthermore, we consider the penalty functions in the presence of short sales (gray area) and no short sales (white area).

In Figure 2, we can see that the LASSO (shown in black) is only effective when short sales are permitted, while the presence of the budget constraint makes it ineffective in the long-only area. In contrast, the RIDGE attains its minimum for an equally weighted portfolio, and when short sales are restricted. Similarly, the SLOPE penalty (shown in red) also reaches its minimum at the equally weighted solution (i.e., $w_1 = w_2 = 0.5$). Still, from a financial perspective, we prefer SLOPE over the RIDGE estimator, because it can promote sparsity by exploiting the singularities.

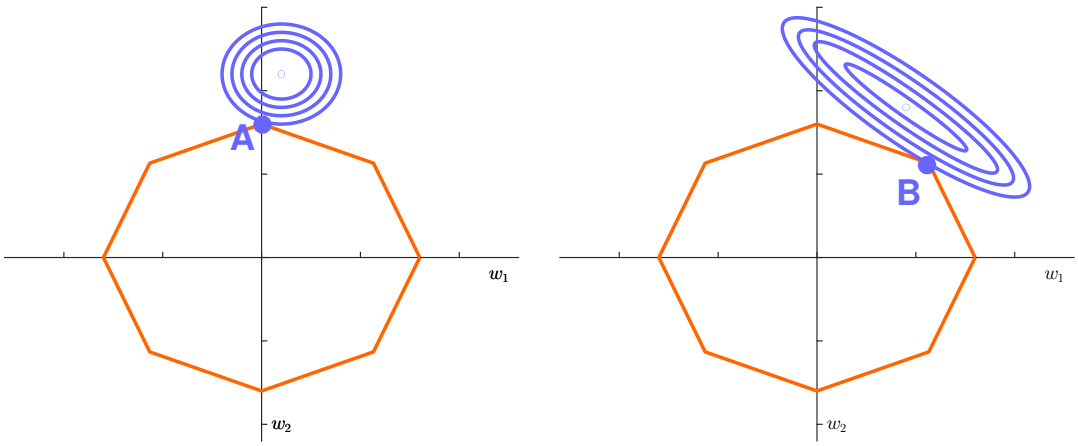
Figure 3 plots the contours of the objective function, together with those of the SLOPE spheres for the two asset case, and when we do not impose a budget constraint (i.e. $\sum_{i=1}^{k} w_i = 1$), as well as considering orthogonal and correlated designs. As noted before, if only $\lambda_2 > 0$, SLOPE always has singular points when one of the asset weights is equal to zero, thereby promoting sparsity. When $\lambda_1 > \lambda_2 > 0$, that is, the sequence is monotonically decreasing, then SLOPE has additional singular points, which correspond to $|w_1| = |w_2|$. This is an appealing property in the presence of correlated data. Specifically, as Panel (b) shows, strong correlation between assets lead to the same weights and thereby grouping. This is consistent with portfolio theory, as it is known that, if assets have all the same correlation coefficients, as well as identical means and variances, the EW is the unique optimal portfolio. SLOPE then allows us to automatically group assets with similar properties.

Figure 2: Penalty Functions in a Two Asset Universe with Budget Constraint




The figure plots the SLOPE coefficient alongside the LASSO ($\ell_1 - Norm$) and the RIDGE penalty ($\ell_2 - Norm$), for a two asset case and under the condition that $w_1 + w_2 = 1$.

Figure 3: Sorted $\ell_1$-Norm Penalty without Budget Constraint




(a) (b)

The figure plots in Panel (a) and (b), the Sorted $\ell_1$-Norm Penalty (SLOPE) in a 2-dimensional setting, considering orthogonal design and correlated design, respectively.

## 2.2 Optimization Algorithm

In this section, we introduce the details of our solution algorithm, which is based on equivalent reformulations of the Alternating Direction Method of Multipliers (ADMM) approach. Our algorithm can also be used to solve the LASSO optimization problem, which is a specific instance of SLOPE. In Section 6.1 in the Appendix, we provide a direct comparison of our algorithm to the state-of-art Cyclic Coordinate Descent (CyCoDe) for LASSO, considering a simulated constant correlation model.

**Reformulation and an ADMM Algorithm.** In order to facilitate the application of proximal operators involving $\rho_\lambda$, we first reformulate (2) - (3) into the following form:

$$\begin{aligned} \min_{\boldsymbol{w}\in\mathbb{R}^k} \quad & \frac{\phi}{2}\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} - \boldsymbol{\mu}'\boldsymbol{w} + \rho_\lambda(\boldsymbol{v}) \\ \text{s.t.} \quad & \boldsymbol{w} = \boldsymbol{v}, \\ & \sum_{i=1}^{k} w_i = 1 \quad , \end{aligned} \tag{4}$$

where $\rho_\lambda(\boldsymbol{w}) := \sum_{i=1}^{k} \lambda_i |w|_{(i)}$ is the sorted $\ell_1$-Norm corresponding to the sequence $\boldsymbol{\lambda}_{SLOPE} = (\lambda_1, \ldots, \lambda_k)'$ satisfying $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_k \geq 0$. To solve (4) we design an ADMM algorithm, which is a variant of the augmented Lagrangian scheme that uses partial updates for the dual variables (for detailed discussion of ADMM see e.g. Boyd et al. (2011)). In our case the associated augmented Lagrangian is given as:

$$\begin{aligned} \mathcal{L}_\eta(\boldsymbol{w}, \boldsymbol{v}; \boldsymbol{\alpha}, \beta) = \frac{\phi}{2}\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} - \boldsymbol{\mu}'\boldsymbol{w} + \rho_\lambda(\boldsymbol{v}) + \boldsymbol{\alpha}'(\boldsymbol{w} - \boldsymbol{v}) + \beta(\boldsymbol{e}'\boldsymbol{w} - 1) \\ + \frac{\eta}{2}\left\{\|\boldsymbol{w} - \boldsymbol{v}\|^2 + (\boldsymbol{e}'\boldsymbol{w} - 1)^2\right\}, \end{aligned} \tag{5}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^k, \beta \in \mathbb{R}$, $\boldsymbol{e}_{k\times 1} = (1, ..., 1)'$, $\boldsymbol{I}_{k\times k}$ is the identity matrix, and $\eta > 0$ is a penalty parameter. The ADMM algorithm consists of the updates:

$$\begin{cases} w^{j+1} & = \arg\min_{\boldsymbol{w}} \ \frac{\phi}{2}\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} - (\boldsymbol{\mu} - \boldsymbol{\alpha}^j - \beta^j\boldsymbol{e})'\boldsymbol{w} + \frac{\eta}{2}(\|\boldsymbol{w} - \boldsymbol{v}^j\|^2 + (\boldsymbol{e}'\boldsymbol{w} - 1)^2) \\ & = \arg\min_{\boldsymbol{w}} \ \frac{1}{2}\boldsymbol{w}'(\phi\boldsymbol{\Sigma} + \eta(I + \boldsymbol{e}\boldsymbol{e}'))\boldsymbol{w} - (\boldsymbol{\mu} - \boldsymbol{\alpha}^j - \beta^j\boldsymbol{e} + \eta(\boldsymbol{v}^j + \boldsymbol{e}))'\boldsymbol{w} \\ & = (\phi\boldsymbol{\Sigma} + \eta(I + \boldsymbol{e}\boldsymbol{e}'))^{-1}(\boldsymbol{\mu} - \boldsymbol{\alpha}^j - \beta^j\boldsymbol{e} + \eta(\boldsymbol{v}^j + \boldsymbol{e})) \\ v^{j+1} & = \arg\min_{\boldsymbol{v}} \ \frac{\eta}{2}\|\boldsymbol{v} - w^{j+1} - 1/\eta\alpha^j\|^2 + \rho_\lambda(\boldsymbol{v}) \\ & = \text{prox}_{\rho_{\lambda/\eta}}(w^{j+1} + (1/\eta)\alpha^j) \\ \alpha^{j+1} & = \alpha^j + \eta(w^{j+1} - v^{j+1}) \\ \beta^{j+1} & = \beta^j + \eta(\boldsymbol{e}'w^{j+1} - 1) \ , \end{cases} \tag{6}$$

where $\text{prox}_{\rho_{\lambda/\eta}}$ is the proximal operator of the Sorted $\ell_1$-Norm corresponding to the sequence $\lambda/\eta$, provided e.g. in Bogdan et al. (2013, 2015).

**A Dual Formulation and the Primal-Dual Gap.** The stopping criterion for our algorithm is based on the Primal-Dual Gap, which we estimate using the following approach. First, taking the infimum over $(\boldsymbol{w}, \boldsymbol{v})$ of the Lagrangian, we get the dual objective,

$$\begin{aligned} g(\boldsymbol{\alpha}, \beta) &= \inf_{\boldsymbol{w}} \frac{\phi}{2}\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} - (\boldsymbol{\mu} - \boldsymbol{\alpha} - \beta\boldsymbol{e})'\boldsymbol{w} - \beta + \inf_{\boldsymbol{v}}\{-\boldsymbol{\alpha}'\boldsymbol{v} + \rho_\lambda(\boldsymbol{v})\} \\ &= \inf_{\boldsymbol{w}} \frac{\phi}{2}\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} - (\boldsymbol{\mu} - \boldsymbol{\alpha} - \beta\boldsymbol{e})'\boldsymbol{w} - \beta - \rho^*_\lambda(\boldsymbol{\alpha}). \end{aligned} \tag{7}$$

From the optimality condition for the infimum over $\boldsymbol{w}$, we have

$$w^* = \phi^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\alpha} - \beta\boldsymbol{e}). \tag{8}$$

Also,

$$\rho^*_\lambda(\boldsymbol{\alpha}) = \sup_{\boldsymbol{v}}\{\boldsymbol{\alpha}^T\boldsymbol{v} - \rho_\lambda(\boldsymbol{v})\} = \begin{cases} 0 & \text{if } \boldsymbol{\alpha} \in C_\lambda \\ +\infty & \text{o.w.} \end{cases} \tag{9}$$

where $C_\lambda := \{\boldsymbol{v} : \mathbb{R}^k : \rho^D_\lambda(\boldsymbol{v}) \leq 1\}$ is the unit sphere defined in the dual norm $\rho^D_\lambda(\cdot)$ of $\rho_\lambda(\cdot)$. Plugging-in these, we get the dual problem

$$\begin{aligned} \max_{\alpha,\beta} \quad & -\frac{1}{2\phi}(\boldsymbol{\mu} - \boldsymbol{\alpha} - \beta\boldsymbol{e})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\alpha} - \beta\boldsymbol{e}) - \beta \\ \text{s.t.} \quad & \boldsymbol{\alpha} \in C_\lambda. \end{aligned} \tag{10}$$

We can estimate the primal-dual gap using (8),

$$\mathcal{G}(\boldsymbol{w}^*, \boldsymbol{v}^*, \boldsymbol{\alpha}^*, \beta^*) = -(\boldsymbol{\alpha}^* + \beta^*\boldsymbol{e})'\boldsymbol{w}^* + \beta^* + \rho_\lambda(\boldsymbol{v}^*) \tag{11}$$

given the dual feasibility of $\boldsymbol{\alpha}^*$, i.e., $\rho^D_\lambda(\boldsymbol{\alpha}^*) \leq 1$.

**Bounds on the Objective Function.** To solve the mean-variance problem, as stated in (2), the investor needs to provide an estimate of the true covariance matrix of asset returns $\boldsymbol{\Sigma}$ and of the true mean $\boldsymbol{\mu}$, which are in the most simplest form given by the sample covariance matrix $\hat{\boldsymbol{\Sigma}}$ and the sample mean $\hat{\boldsymbol{\mu}}$, respectively. However, $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\mu}}$ might be prone to substantial estimation errors and highly sensitive to outliers. Let us define $M(\boldsymbol{\Sigma}, \boldsymbol{\mu}) = \frac{\phi}{2}\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} - \boldsymbol{w}'\boldsymbol{\mu}$, where $\boldsymbol{w}$ is the vector of weights returned by SLOPE. Now, observe that the Sorted $\ell_1$-Norm satisfies $\rho_\lambda(\boldsymbol{w}) \geq \lambda_k\|\boldsymbol{w}\|_1$. Thus, as $\lambda_k > 0$, we have $\|\boldsymbol{w}\|_1 \leq c$, with $c = \frac{\rho_\lambda(\boldsymbol{w})}{\lambda_k}$, and simple calculations following the results of Fan et al.

(2012) for LASSO, yield:

$$|M(\widehat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\mu}}) - M(\boldsymbol{\Sigma}, \boldsymbol{\mu})| \leq \frac{\phi}{2} ||\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}||_\infty \rho_\lambda^2(\boldsymbol{w}) / \lambda_k^2 + ||\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}||_\infty \rho_\lambda(\boldsymbol{w}) / \lambda_k \tag{12}$$

where $||\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}||_\infty$ and $||\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}||_\infty$ are the maximum component-wise estimation errors for the covariance matrix and the expected return. This result implies that the difference between the objective functions for the estimated and true vector of parameters decreases, as we restrict the Sorted $\ell_1$-Norm of the weight vector.

It is also important to observe that, due to imposing the budget constraint, a higher weight on the penalty sets an upper bound on the total amount of short sales in the portfolio, as $\rho_\lambda(\mathbf{w}) \geq \lambda_{\mathbf{k}} ||\boldsymbol{w}||_{\mathbf{1}} = \lambda_{\mathbf{k}}(\boldsymbol{w}^+ + \boldsymbol{w}^-)$, with $\boldsymbol{w}^+ - \boldsymbol{w}^- = 1$, where $\boldsymbol{w}^+ = \sum_{w_i \geq 0} w_i$ and $\boldsymbol{w}^- = \sum_{w_i < 0} w_i$ are the gross amount of long and short positions, respectively.

# 3 Simulation Analysis

In this section, we analyze and explain the effect of SLOPE on the model risk, the sparsity and the grouping properties, by considering simulated data. Assuming that $\boldsymbol{\Sigma}$ is known, we can use the alternative formulation of SLOPE and define:

$$\boldsymbol{w}_{opt} = \underset{\boldsymbol{w}: \sum_{i=1}^k w_i = 1,\, \rho_{\lambda(\boldsymbol{w})} \leq c}{\arg\min} \frac{\phi}{2} \boldsymbol{w}' \boldsymbol{\Sigma} \boldsymbol{w} - \boldsymbol{\mu}' \boldsymbol{w} \tag{13}$$

$$\hat{\boldsymbol{w}}_{opt} = \underset{\boldsymbol{w}: \sum_{i=1}^k w_i = 1,\, \rho_{\lambda(\boldsymbol{w})} \leq c}{\arg\min} \frac{\phi}{2} \boldsymbol{w}' \hat{\boldsymbol{\Sigma}} \boldsymbol{w} - \hat{\boldsymbol{\mu}}' \boldsymbol{w} \tag{14}$$

whereas $\boldsymbol{w}_{opt}$ and $\hat{\boldsymbol{w}}_{opt}$ are the theoretical optimal and the empirical optimal weights vector, respectively. We then define the *empirical* portfolio risk as $\widehat{Risk}(\hat{\boldsymbol{w}}_{opt}) = \hat{\boldsymbol{w}}'_{opt} \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{w}}_{opt}$,

the *actual* portfolio risk as $Risk(\hat{\boldsymbol{w}}_{opt}) = \hat{\boldsymbol{w}}'_{opt}\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_{opt}$ and the *oracle* portfolio risk as $Risk(\boldsymbol{w}_{opt}) = \boldsymbol{w}'_{opt}\boldsymbol{\Sigma}\boldsymbol{w}_{opt}$, respectively. Following the proof of Theorem 1 of Fan et al. (2012), we can easily show that in case when $\lambda_k > 0$, the pair differences between the three measures are upper bounded by:

$$|Risk(\hat{\boldsymbol{w}}_{opt}) - Risk(\boldsymbol{w}_{opt})| \leq 2c^2\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_\infty, \tag{15}$$

$$|Risk(\hat{\boldsymbol{w}}_{opt}) - \widehat{Risk}(\hat{\boldsymbol{w}}_{opt})| \leq c^2\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_\infty, \tag{16}$$

$$|Risk(\boldsymbol{w}_{opt}) - \widehat{Risk}(\hat{\boldsymbol{w}}_{opt})| \leq c^2\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_\infty \tag{17}$$

The three risk measures then allow us to extract different information: The empirical risk is the only one that is known in a practical setting and is estimated from our in-sample data. The actual risk is the one, to which the investor is truly exposed to, when setting up a portfolio with optimal weights ($\hat{\boldsymbol{w}}_{opt}$). Finally, the oracle risk is the risk the investor could only obtain if $\boldsymbol{\Sigma}$ is known. As the SLOPE penalty becomes more binding when $\boldsymbol{\lambda}$ "increases", the three risk measures get closer to each other. In the following sections, we consider two simulation set-ups, and show how increasing the SLOPE penalty allows to reduce the estimation error and to avoid its accumulation in the portfolio risk.

**Hidden Factor Structure.** Let us assume that the return of an asset is represented by a linear combination of $r$ risk factors. Furthermore, let $t$ be the number of observations, $k$ be the number of assets, and $\boldsymbol{F}_{t\times r} = [\boldsymbol{f}_1\ \boldsymbol{f}_2\ ...\ \boldsymbol{f}_r]$, where $\boldsymbol{f}_i$ is the $t \times 1$ vector of returns of the $i^{th}$ risk factor. Moreover, let $\boldsymbol{B}_{r\times k}$ be the loading matrix for the individual risk factors. Then, the $t \times k$ matrix of asset returns from the Hidden Factor Model (i.e. $\boldsymbol{R}_{HF}$)

can be represented as:

$$\boldsymbol{R}_{HF} = \boldsymbol{F} \times \boldsymbol{B} + \boldsymbol{\epsilon} \tag{18}$$

where $\boldsymbol{\epsilon}$ is a $t \times k$ matrix of error terms.

For our first illustration of the performance of SLOPE, we generate the data using the following simplified scenario:

- $t = 50$, $k = 12$, $r = 3$,
- the risk factors $f_1, \ldots, f_3$ are independent from the multivariate standard normal $N(0, I_{r\times r})$ distribution, with $I_{r\times r}$ being an identity matrix,
- the vectors of error terms $\epsilon_i$, $i = 1, \ldots, k$, for each asset are independent from each other, as well as from each of the risk factors and come from the multivariate normal distribution $N(0, 0.05 \times I_{r\times r})$
- the loadings matrix $B_{r\times k}$ is made of exactly four copies of each of the following columns: $[0.77\ 0.64\ \ 0]'$, $[0.9\ \ 0\ \ -0.42]'$ and $[0\ \ 0.31\ 0.64]'$.

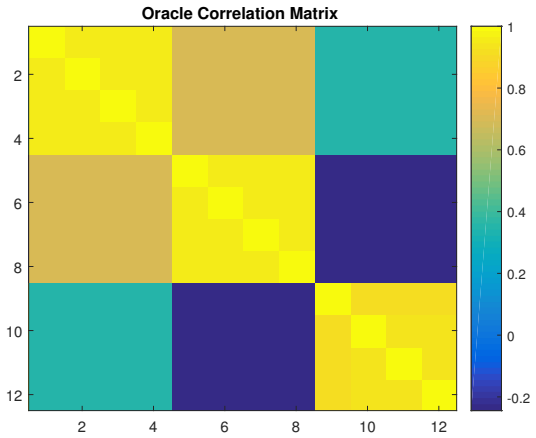
In this way, we generate three different groups that have the same exposure to the same two risk factors and are thus strongly correlated.[3]

Finally, given (18), the covariance matrix of the assets $\boldsymbol{\Sigma}_{HF}$ is given by:

$$\boldsymbol{\Sigma}_{HF} = \boldsymbol{B}'\boldsymbol{B} + 0.05 \times I_{k\times k}. \tag{19}$$

[3]For the robustness of our results, we tested SLOPE in various set-ups, with qualitatively similar results, but restrict ourselves, due to space limitations, to the most interesting one. The results of the remaining simulations are available from the authors upon request.

Figure 4: Hidden Factors Correlation Matrix




The figure plots the correlation matrix, based on the modeled Hidden Factor Structure, considering a universe of 12 assets, of which 4 are always exposed to exactly two out of the three risk factors in the market.

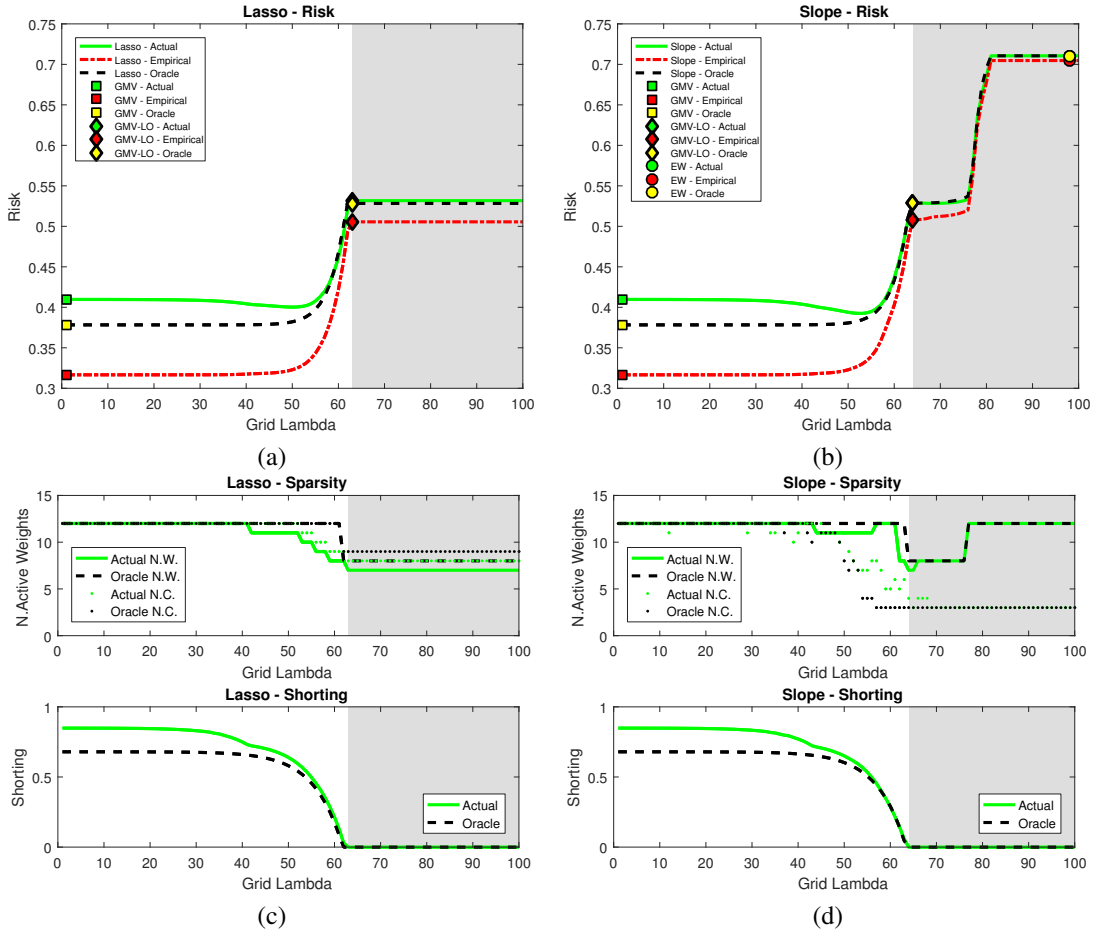
After generating our $t \times k$ matrix $\boldsymbol{R}_{HF}$ of asset returns from (18), we can then estimate $\boldsymbol{\Sigma}_{HF}$, using the sample covariance estimate $\hat{\boldsymbol{\Sigma}}_{HF}$. Figure 4 shows the correlation matrix resulting from (19) and assuming that the four assets from each of the three groups are clustered together. The figure illustrates that our simulation scenario explicitly models a block correlation environment, with strong correlation among each of the assets having the same underlying risk factor exposures, and low to negative correlations between the assets with a different underlying factor structure. Subsequently, we first consider a minimum variance optimization, such that $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu} = 0$. In a second step, we solve the mean-variance problem, estimating $\boldsymbol{\mu}$ by the sample average of the asset's returns. In both setups, we investigate the behavior of SLOPE and the LASSO with respect to portfolio risk, and given that we vary the value of the tuning parameter.

Unlike the LASSO or the RIDGE penalties, SLOPE requires us to define a decreasing sequence of $\boldsymbol{\lambda}_{SLOPE} = (\lambda_1, \lambda_2, \ldots, \lambda_k)$. For our analysis, we use the decreasing sequence of quantiles of the standard normal distribution, as in Bogdan et al. (2013) and Bogdan

et al. (2015), with $\lambda_i = \alpha\Phi^{-1}(1 - q_i)$, $\forall i = 1, ..., k$, where $\Phi$ is the cumulative distribution function of the standard normal distribution and $q_i = i \times \theta/2k$, and in which $\theta = 0.01$, regulates how fast the sequence of lambda parameters is decreasing. In our simulations, we vary the scaling parameter $\alpha$ so that the first element of the sequence $\lambda_1 = \alpha\Phi^{-1}(1 - q_1)$ is equal to a grid of 100 log-spaced values between $10^{-5}$ and $10^2$. Note that in the case of the LASSO, we only choose one lambda parameter, which then remains constant for all assets. In our simulation and also in the real world analysis, we always choose $\lambda_{LASSO} = \lambda_1$. This choice favors sparser solutions for the LASSO, since for the remaining $k - 1$ assets its penalty is larger than that of SLOPE.

Figure 5 plots the resulting risk and weight profile for the minimum variance optimization, when we solve (2) separately with the LASSO and the SLOPE penalties for the grid of 100 lambda parameters, and considering $\mathbf{\Sigma}_{HF}$ and the sample covariance estimate $\widehat{\mathbf{\Sigma}}_{HF}$, respectively. In particular, Panels (a) and (b) show the risk profile of the LASSO and SLOPE, i.e. the actual, the oracle, and the empirical risk, together with the results of the GMV, the GMV-LO and the EW portfolios. For both, the oracle and the actual solution, Panels (c) and (d) display on top the number of active weights together with the number of groups, that is the number of distinct coefficients, while on the bottom, it shows the amount of shorting (i.e. $\boldsymbol{w}^-$). The grey surface indicates the no-short-sale-area (i.e. $w_i \geq 0 \; \forall \; i = 1, .., k$). Figure 5 shows that for a tuning parameter equal to zero, which corresponds to the GMV solution, the empirical risk is about 1.3 times lower than the actual risk (Panels (a) and (b)), with 12 active positions (Panel (c)) and slightly under 100% short sales (Panel (d)). This can be interpreted as evidence that in over-fitted models the estimation error in $\widehat{\mathbf{\Sigma}}_{HF}$ strongly affects the estimation of the asset weights. As here neither the LASSO nor the SLOPE penalty are binding, estimation errors can enter unhindered into the optimization. Michaud (1989) describes this phenomenon as

Figure 5: Hidden Factors Minimum-Variance Profile




The figure shows the Hidden Factor minimum-variance risk profile for the LASSO and the SLOPE, including in Panel (a) and (b) their actual, empirical and oracle risk profiles, together with that of the GMV, the GMV-LO and the EW solutions. Furthermore, Panel (c) and (d) display the number of active weights, together with the grouping profile (top) and the total amount of shorting (bottom). All values are computed based on a Hidden Factor Structure, with three risk factors and considering for the exponentially decreasing sequence of lambda parameters, a grid of 100 log spaced starting points for $\lambda_1$ from $10^{-5}$ (i.e. x-value = 1) to $10^{2}$ (i.e. x-value = 100).

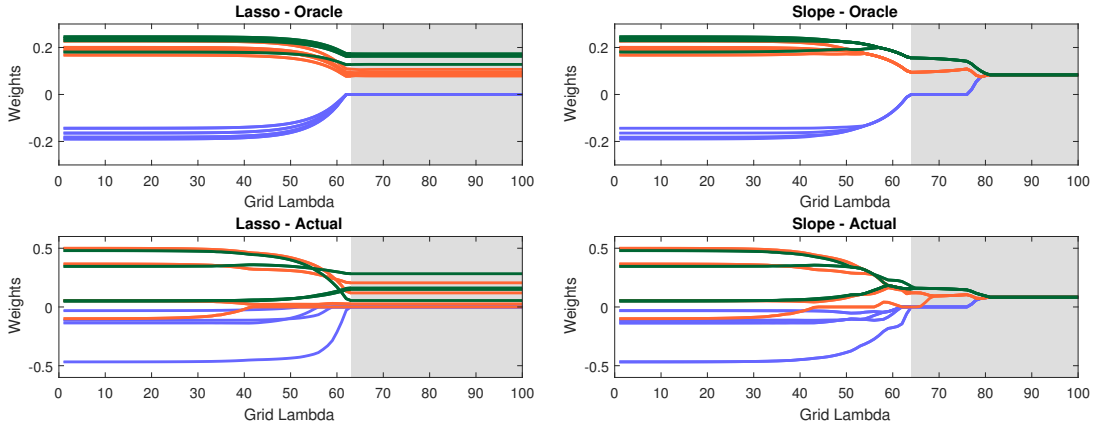
"error maximization", in which the ill-conditioned covariance estimates are amplified through the optimization, leading to extreme long and short portfolio weights. Moving along the grid of $\boldsymbol{\lambda}$ parameters from the left to the right, Panels (c) and (d) show that the two penalties reduce the total amount of shorting in the oracle and the actual portfolio. As we move from the GMV towards the GMV-LO, the actual, oracle, and empirical risk

of the LASSO and the SLOPE align. This effect was first observed and theoretically motivated by Fan et al. (2012), showing that the portfolio risk evolves in a U-shape, in which risk first decreases before increasing again, due to the restriction of short sales. With the observations above, we extend the results of Fan et al. (2012), showing that the U-shaped behavior of the portfolio risk is not the only possible one. Especially when the dependence among the assets is positive, the tighter constraint in terms of short sales shrinks the optimization search space of feasible solutions, making it impossible to exploit the optimal diversification benefits. This leads to a higher portfolio risk when reaching the GMV-LO. The investor also reaches the maximum sparsity, that is the maximum number of coefficients equal to zero, at this point. For the LASSO, increasing the tuning parameter beyond this point does not alter the allocation any further, as the regularization penalty is constant and equal to 1.

This is different for SLOPE: in fact, Figure 6 shows the evolution of the portfolio weights for both the oracle and the actual solution, considering both the LASSO and the SLOPE penalty. As before, the grey surface indicates the no-short-sale-area.

From Figure 6, we can observe two important characteristics of SLOPE: First, while the LASSO shrinks the weights up until the no short sale area, all non-zero coefficients still receive a different weight, independent of their underlying factor exposures. SLOPE, on the other hand, is able to identify the three distinct types of securities, consistent with the true model, and groups them together, by assigning the same coefficient values to them. This provides information about the dependence structure among the assets, and gives the investor the flexibility to select from the groups the assets, which best fit her individual preferences. Not surprisingly, the oracle risk starts to form groups among the securities even before entering into the no short sale area, while the actual weights can only capture the underlying structure much later, and when we already impose a larger

Figure 6: Hidden Factors Minimum-Variance Weight Profiles



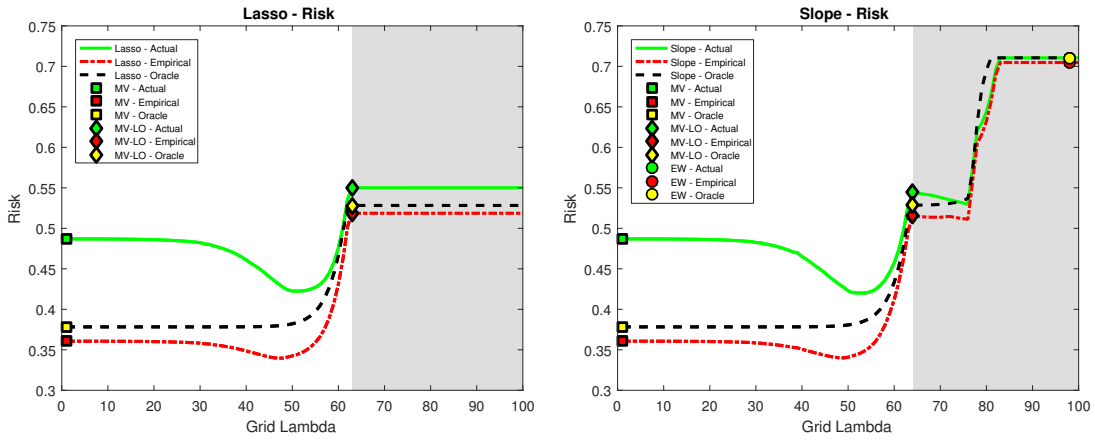

(a) (b)

The figure shows the weight profile of the oracle (top) and actual (bottom) solution of the LASSO and the SLOPE penalty, considering a minimum variance setup. All values are computed based on a Hidden Factor Structure, with three risk factors and considering for the exponentially decreasing sequence of lambda parameters, a grid of 100 log spaced starting points for $\lambda_1$ from $10^{-5}$ (i.e. x-value = 1) to $10^2$ (i.e. x-value = 100). Equally colored weights characterize assets with the same underlying factor exposure.

tuning parameter value. Second, and different to the LASSO, increasing the lambda parameters past the point of the GMV-LO, the octagonal shape of the penalty pushes the solution towards the equally weighted portfolio. That is, the aforementioned grouping effect increases, and all weights - even those that were shrunken towards zero - are assigned the same coefficient value of $\frac{1}{k}$. Given that the equally weighted portfolio is only optimal when all assets have the same risk and return characteristics, in our example, this allocation results in higher portfolio risk when compared to the GMV-LO or GMV portfolios.

In a next step, we investigate how SLOPE performs in a mean-variance framework. For that purpose, we choose $\boldsymbol{\mu}_{HF} = 0$ and $\hat{\boldsymbol{\mu}}_{HF} = mean(\boldsymbol{R}_{HF})$ as the sample average of the $t = 50$ return observations for the $k = 12$ assets. Figure 7 plots the resulting risk profile of the LASSO and the SLOPE, when we again consider 100 logspaced values from $10^{-5}$ to $10^2$ for the starting point of the sequence of tuning parameters. The effect of adding the mean estimate to the minimum variance framework can be observed from multiple

Figure 7: Hidden Factors Mean-Variance Profile




(a) (b)

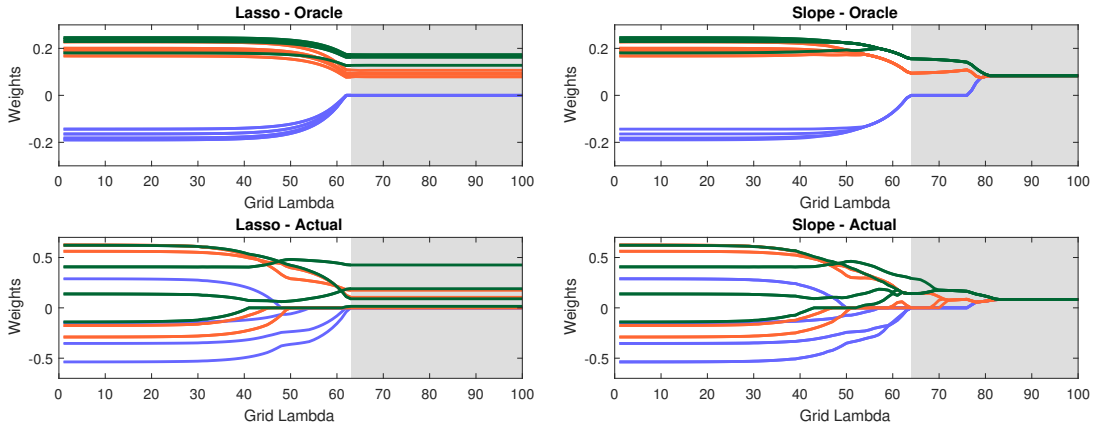
The figure shows the Hidden Factor mean-variance risk profile for the LASSO and the SLOPE, including in Panel (a) and (b) their actual, empirical and oracle risk profiles together with that of the MV, MV-LO and EW solutions. All values are computed based on a Hidden Factor Structure, with three risk factors and considering for the exponentially decreasing sequence of lambda parameters, a grid of 100 log spaced starting points for $\lambda_1$ from $10^{-5}$ (i.e. x-value = 1) to $10^{2}$ (i.e. x-value = 100).

viewpoints: First, from the risk plot in Panels (a) and (b), we observe that the estimation errors, which are known to be larger in the mean than in the covariance matrix (see for example Merton (1980), Michaud (1989), DeMiguel et al. (2009a)), lead to a more pronounced difference among the empirical, oracle, and actual risk. The unconstrained mean-variance solution (MV) now has an empirical risk of 0.35, compared to the actual risk of 0.5. Increasing the weight on the tuning parameter, for both the LASSO and the SLOPE, reduces the estimation error, letting the three risk measure converge, as we move from the MV to the mean-variance long-only solution (MV-LO).

Finally, Figure 8 shows that, regardless of the penalty function, the oracle and actual weight vectors differ substantially in their values. As before, weights for the LASSO, of assets with different underlying factor exposures, are overlapping, and randomly picked. SLOPE, on the other hand, and despite requiring larger $\lambda$ values, is still capable to correctly group the assets with the same underlying risk factor exposure, and thereby dis-

entangling signal from noise.

Figure 8: Hidden Factors Mean-Variance Weight Profiles




(a) (b)

The figure shows the weight profile of the oracle (top) and actual (bottom) solution for both the LASSO and SLOPE penalty, considering a mean-variance setup. All values are computed based on a Hidden Factor Structure, with three risk factors and considering for the exponentially decreasing sequence of lambda parameters, a grid of 100 log spaced values for $\lambda_1$ from $10^{-5}$ (i.e. x-value = 1) to $10^2$ (i.e. x-value = 100). Different colors characterize assets with the same underlying factor exposure.

**SP500 Simulated Covariance Matrix.** Our generic example suffers from the drawback that, in reality, assets do not follow such a strict exposure to only two out of the three underlying risk factors, but are most likely exposed to all factors. Furthermore, investors often face an investment universe that is larger than 12 assets.

To model a more realistic scenario, we consider a factor setup, which was introduced by Fan et al. (2008) and which is driven by the estimated covariance matrix between different assets. As the behavior of the LASSO in a high dimensional environment is widely studied, and for the sake of brevity, we restrict ourselves to study the behavior of the new SLOPE procedure. The results for the performance of the LASSO in such an environment are available from the authors upon request.

As before, we assume that security returns can be expressed as a linear combination of

risk factors, as in (18), implying that our true covariance matrix takes the form

$$\boldsymbol{\Sigma}_{\text{SP500}} = \boldsymbol{B}'\boldsymbol{\Sigma}_F\boldsymbol{B} + \boldsymbol{\Sigma}_\epsilon \ , \tag{20}$$

where $\boldsymbol{\Sigma}_F$ is the covariance matrix of hidden factors and $\boldsymbol{\Sigma}_\epsilon$ is the covariance matrix of error terms.

Different to the simulations before, we draw the three factors and the factor loadings from multivariate normal distributions that are calibrated to real world data. Specifically, our three hidden factors are generated from a trivariate normal distribution $N(\boldsymbol{\mu}_F, \boldsymbol{\Sigma}_F)$, where $\boldsymbol{\mu}_F$ and $\boldsymbol{\Sigma}_F$ are calculated based on data taken from Kenneth French's Homepage[4], spanning the time period from 31.12.2004 to 31.01.2016 (see Table 1). The factor loadings for each of $k = 500$ assets are independently drawn from the trivariate normal distribution $N(\boldsymbol{\mu}_B, \boldsymbol{\Sigma}_B)$, in which $\boldsymbol{\mu}_B$ and $\boldsymbol{\Sigma}_B$ are calculated using data on the SP500 from 31.12.2004 to 31.01.2016, as reported in Table 1. Finally, the idiosyncratic noises are generated from a gamma distribution with shape parameter $\alpha = 7.2609$ and scale parameter $\beta = 0.0028$ (see Fan et al. (2008) for details).

We then set $t = 500$ and solve (2) both with (a) $\hat{\boldsymbol{\mu}}_{\text{SP500}} = \boldsymbol{\mu}_{\text{SP500}} = 0$, which is then the minimum variance problem, and with (b) $\boldsymbol{\mu}_{\text{SP500}} = \boldsymbol{B}'\boldsymbol{\mu}_F$, in which $\hat{\boldsymbol{\mu}}_{SP500}$ is equal to the sample mean of the last 500 return observations obtained from (18). We use the same $\boldsymbol{\lambda}_{SLOPE}$ sequences as in the previous section, with the first element of the sequence $\lambda_1$ taking 100 log spaced values from $10^{-8}$ to $10^{-1.5}$ in the minimum variance framework, and from $10^{-4}$ to $10^{-1.5}$ for the mean-variance framework.

Figure 9 shows the risk and sparsity profiles for the minimum variance optimization, while Figure 10 displays the results for the mean-variance set-up. Both plots are consis-

[4] `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

Table 1: Configuration Parameters for Fan et al. (2008) Simulation

| Parameters for factor loadings | | | | Distribution of hidden factors | | | |
|---|---|---|---|---|---|---|---|
| $\boldsymbol{\mu}_B$ | $\boldsymbol{\Sigma}_B$ | | | $\boldsymbol{\mu}_F$ | $\boldsymbol{\Sigma}_F$ | | |
| -0.0679 | 0.0062 | -0.0016 | 0.0020 | 0.00022 | 0.000157 | 0.000015 | 0.000028 |
| 0.1505 | -0.0016 | 0.0109 | 0.0012 | 0.00012 | 0.000015 | 0.000033 | -0.000016 |
| -0.0203 | 0.0020 | 0.0012 | 0.0173 | -0.00018 | 0.000028 | -0.000016 | 0.000034 |

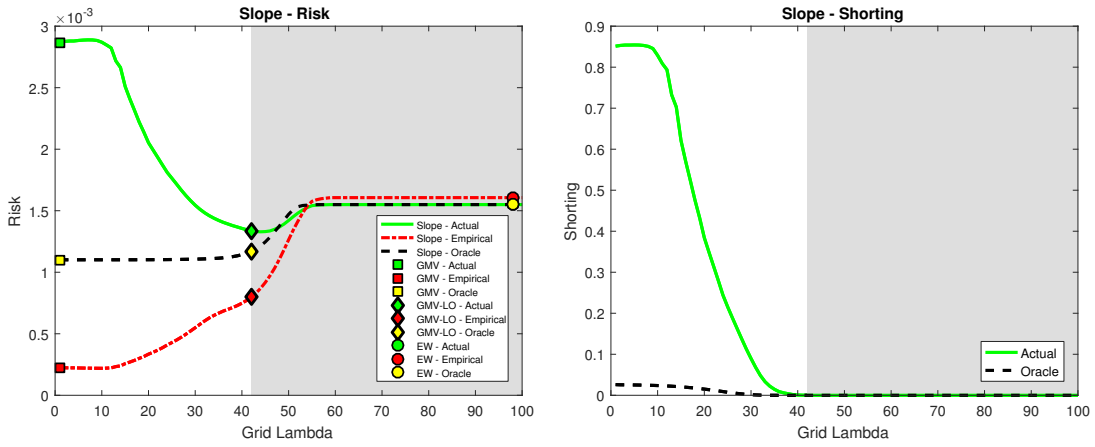
The table reports the means ($\boldsymbol{\mu}$) and the covariance matrices ($\boldsymbol{\Sigma}$) used as input parameters for the trivariate normal distributions to sample the hidden factors and the assets' factor loadings. The first and second moments for the factor loadings are calibrated using daily data from the SP500 from 31.12.2004 to 31.01.2016, while the parameters for the factor distribution are calibrated, using daily data for the three Fama-French risk factors plus the Carhart (1997) liquidity factor, covering again the period from 31.12.2004 to 31.01.2016.

tent with the findings from the Hidden Factor Model. In detail: we again observe that in the GMV solution (i.e., $\lambda = 0$), there is a substantial difference among the three risk measures, with the empirical risk (given in red) highly underestimating the actual risk. Still, with a larger $\lambda$, the difference between the risk measures decreases and we move closer to the GMV-LO solution. Furthermore, given a larger investment universe and the realistic dependence structure, the impact of a reduction in the search space, due to a larger penalty is smaller, and the actual risk decreases, as we move closer to the no short selling area and towards the EW solution.

The same is true for the mean-variance set-up. In fact, Figure 10 shows that including the estimate of the mean, $\hat{\boldsymbol{\mu}}_{\text{SP500}}$, introduces as expected even more estimation error, than compared to the minimum variance optimization. Panel (a) of Figure 10 shows that the effect of the mean estimate is so large that even in the EW solution the actual risk is smaller than in the MV solution. However, note that this is different for the oracle solution, which considers $\boldsymbol{\mu}_{\text{SP500}}$ and therefore is not prone to any extreme mean estimates. Still, as we increase the $\boldsymbol{\lambda}$ sequence, the three risk measures align and the overall risk reduces. Most importantly, while the LASSO is not effective in the no-short sale area, SLOPE provides this risk reduction effect even past the MV-LO, and when

moving towards the EW solution.

Figure 9: SP500 Minimum-Variance Profile




(a) (b)

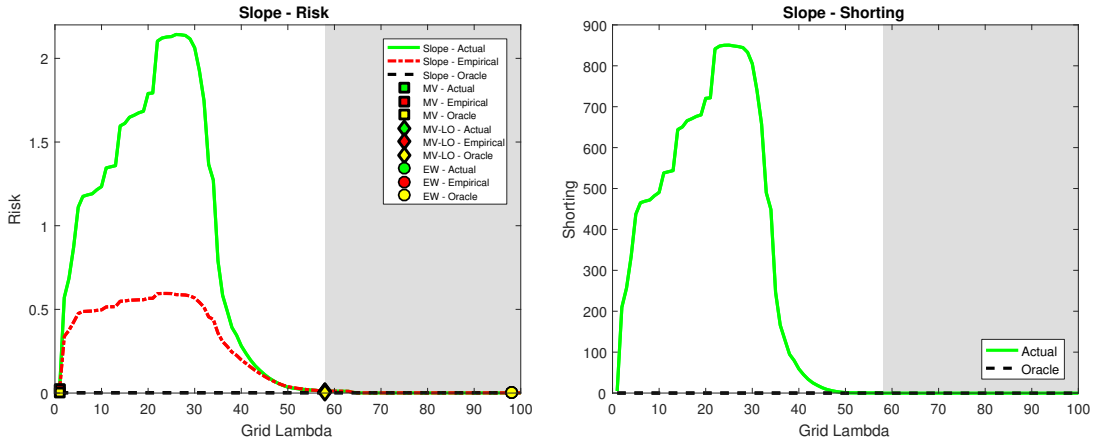
The figure shows the Fan Simulated SP500 minimum-variance risk profile for SLOPE, including in Panel (a) the actual, empirical and oracle risk profile of SLOPE together with that of the GMV, the GMV-LO and the EW solutions, as well as in Panel (b) the total amount of shorting of the actual and oracle solution. All values are computed based on the Fan et al. (2008) simulation, calibrated to the SP500, and considering for the exponentially decreasing sequence of lambda parameters, a grid of 100 log spaced values with a starting point for $\lambda_1$, from $10^{-8}$ (i.e. x-value = 1) to $10^{-1.5}$ (i.e. x-value = 100).

This is especially valuable for investors, who face short sale constraints and deal with large portfolios and many candidate assets. Indeed, as previously discussed, the LASSO would be stuck in the GMV-LO (MV-LO) solution, while with SLOPE, investors facing short sale constraints are then able to further reduce their risk. Furthermore, they are even able to set up new strategies and to exploit the grouping property, by selecting among the assets the one that best corresponds to their financial investment objectives.

An interesting difference between the minimum variance and mean-variance optimization is the behavior of the short sales, as shown in Panel (b) of Figure 9 and Figure 10. While in the minimum variance set-up, we start with an initial solution that shows a certain amount of shorting and which is subsequently reduced with an increasing value of $\lambda$, in the mean-variance optimization, using the sample mean estimate, first leads to an increase in the total amount of shorting, before decreasing it again. For the actual risk,

this behavior can be explained by the presence of positively correlated assets, which is typical in financial markets. The optimization then exploits the diversification, by taking at the beginning extreme positive and negative weights. With an increased $\lambda$, the penalty then has an effect on shrinking large weights and reducing the number of active positions. Hence, to exploit diversification benefits, other assets with smaller weights end up in increasing their negative exposure.

Figure 10: SP500 Mean-Variance Profile




(a) (b)

The figure shows the Fan Simulated SP500 mean-variance risk profile for SLOPE, including in Panel (a) the actual, empirical and oracle risk profile of SLOPE together with that of the MV, the MV-LO and the EW solutions, as well as in Panel (b) the total amount of shorting of the actual and oracle solution. All values are computed based on the Fan et al. (2008) simulation, calibrated to the SP500, and considering for the exponentially decreasing sequence of lambda parameters, a grid of 100 log spaced values with a starting point for $\lambda_1$, from $10^{-4}$ (i.e. x-value = 1) to $10^{-1.5}$ (i.e. x-value=100).

Finally, and even though not displayed, we observe that the desired feature of grouping assets together persists, not only in a directly modeled Hidden Factor Structure, but even in a high dimensional scenario, with real world calibrated covariance matrices. This might be valuable for investors who can then set up sophisticated asset allocation strategies, exploiting SLOPE's grouping property, such as SLOPE-MV, which we introduce in Section 4.

# 4 Empirical Analysis

## 4.1 Set up and Data

This section studies the out-of-sample performance of the SLOPE procedure, considering, as typical for most studies, a minimum variance framework (see i.e. Jagannathan and Ma (2003), Brodie et al. (2009), DeMiguel et al. (2009a), Giuzio and Paterlini (2016)). Our analysis compares SLOPE with state-of-the-art portfolio selection methods, such as the EW, the GMV, the GMV-LO, the equal risk contribution (ERC), the RIDGE and the LASSO portfolio. We examine two extensions to our standard SLOPE procedure: (1) SLOPE with an added long-only constraint (SLOPE-LO) and (2) a portfolio in which we utilize SLOPE's selection ability, by first running SLOPE-LO on the whole time period, identifying groups of similar assets and picking out of each group the one with minimum variance. We then roll through the dataset and solve for these active securities the GMV-LO portfolio (SLOPE-MV).

In the following analysis, we consider five data sets, including the monthly log-return observations for the 10- and 30 Industry Portfolios (Ind), as well as the 100 Fama French (FF) portfolios, formed on Size and Book-to-Market, as well as the daily returns of the SP100 and SP500. The monthly portfolio values are taken from Kenneth French's Homepage[5] and span the period from January 1970 to January 2017 ($T = 565$ monthly observations). The daily return data are obtained from Datastream, covering the period from 31.12.2004 to 31.01.2016 ($T = 2890$ daily observations). Table 2 reports the descriptive statistics for the data sets. As shown by the skewness and the kurtosis values, all of them exhibit the typical return time series characteristics, including fat tails and slight asymmetry.

[5] `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/`

Table 2: Descriptive Statistics of the Dataset

| Dataset | $T$ | $k$ | $\hat{\mu}$ | $\hat{\sigma}$ | $\widehat{med}$ | $\widehat{min}$ | $\widehat{max}$ | $\widehat{skew}$ | $\widehat{kurt}$ | *period* | *freq.* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 10Ind | 565 | 10 | 0.099 | 0.043 | 0.012 | -0.211 | 0.156 | -0.476 | 5.077 | 01/1970 - 01/2017 | Monthly |
| 30Ind | 565 | 30 | 0.010 | 0.048 | 0.012 | -0.255 | 0.179 | -0.507 | 5.749 | 01/1970 - 01/2017 | Monthly |
| 100FF | 565 | 100 | 0.011 | 0.052 | 0.015 | -0.262 | 0.241 | -0.551 | 5.600 | 01/1970 - 01/2017 | Monthly |
| SP100 | 2890 | 93 | 0.000 | 0.013 | 0.000 | -0.098 | 0.116 | -0.240 | 14.816 | 12/2004 - 01/2016 | Daily |
| SP500 | 2890 | 443 | 0.000 | 0.014 | 0.000 | -0.107 | 0.109 | -0.418 | 13.234 | 12/2004 - 01/2016 | Daily |

The table reports descriptive summary statistics for the 10 Industry Portfolios, the 30 Industry Portfolios, the 100 Fama French Portfolios, the S&P 100 and the S&P 500, respectively. Reported are for the daily (monthly) data: the number of observations ($T$), the number of constituents($k$), the mean ($\hat{\mu}$), the standard deviation ($\hat{\sigma}$), the median ($\widehat{med}$), the minimum ($\widehat{min}$), the maximum ($\widehat{max}$), the skewness ($\widehat{skew}$), the kurtosis ($\widehat{kurt}$), the period that the data set covers (*period*) and the frequency (*freq.*).

To evaluate our portfolios in an out-of-sampling setting, we rely on a rolling window approach with a window size of $\tau = 120$ monthly observations for the 10Ind, the 30Ind, and the 100FF, as well as $\tau = 500$ daily observations for the SP100 and SP500.[6] All portfolios are re-balanced monthly, discarding the oldest and including the most recent observations, allowing for a total of $t = 445$ ($t = 115$) out-of-sample returns for the monthly (daily) data.

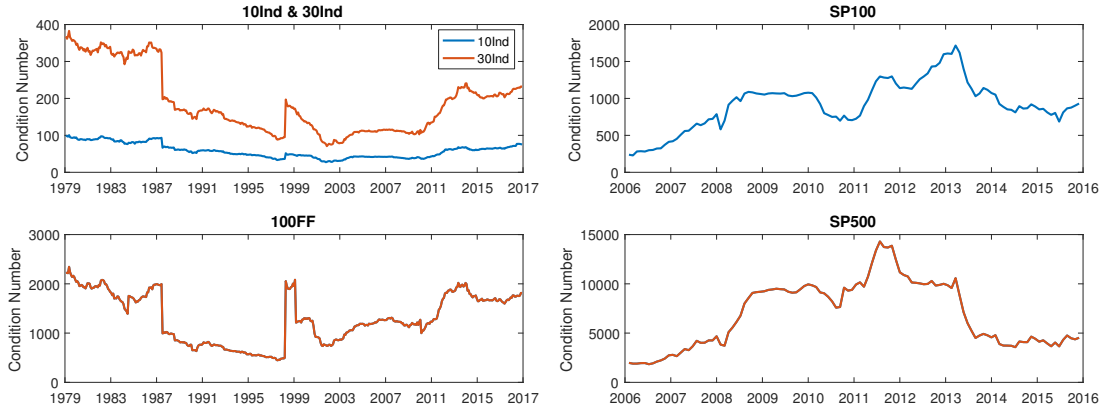
The rolling window approach for the daily data works as follows: the first $\tau$ return observations are used to estimate $\hat{\mathbf{\Sigma}}_t$, according to the shrinkage approach by Ledoit and Wolff (2004). Then, $\hat{\mathbf{\Sigma}}$ is used as the input to compute the optimal weight vector $\hat{\boldsymbol{w}}_t$. The resulting portfolio is assumed to be held for the following 21 days. At $t + 1$, the $k$ constituents' returns over this monthly period, $\boldsymbol{R}_{t+1}$, are used to compute the out-of-sample portfolio return as: $R_{p,t+1} = \hat{\boldsymbol{w}}_t \boldsymbol{R}_{t+1}$. In the next step, we roll the data window forward, dropping the last and adding the most recent 21 observations to our training set. We then estimate a new weight vector, which determines our portfolio holdings and the out-of-sample return for the next month. This process is repeated until the end of

[6] To test the robustness of our results, we account for different window sizes of $\tau = 250, 750$ and 1000 daily observations, and make the results available upon request. The obtained results are qualitative similar.

the data set is reached. The same procedure is applied to the Industry and Fama French portfolios, though the window is rolled forward by one monthly observation instead of 21 daily observations.

Figure 11 plots the condition numbers[7] for the covariance matrix estimate, whereas large values indicate that our estimate is very sensitive to changes in the underlying data structure. These large condition numbers often stem from multicollinearity between the assets and consequently, as we rebalance the portfolio, the changes in the input parameter then lead to extreme changes in the portfolio weights and to high turnover levels.

Figure 11: Condition Numbers

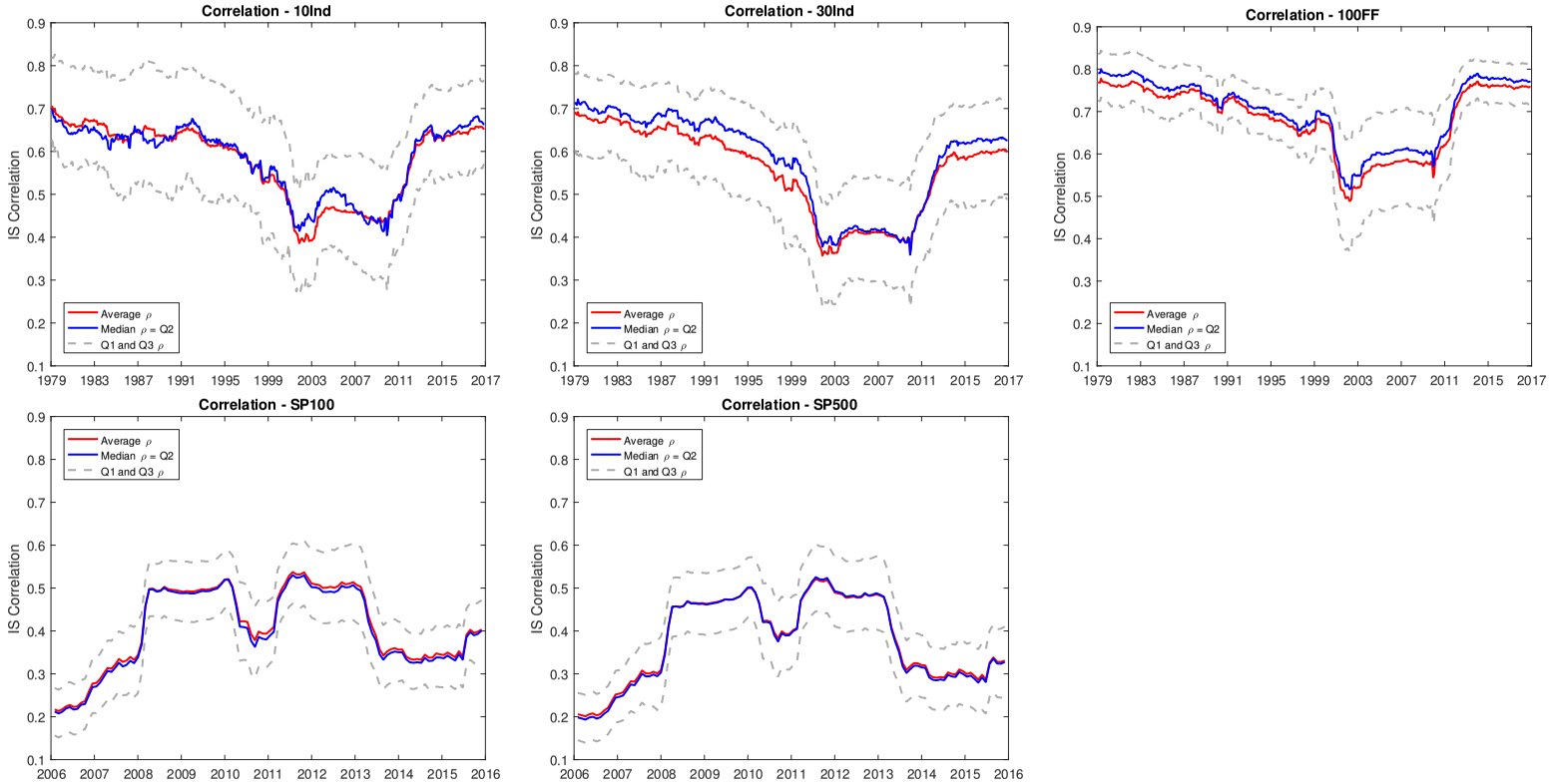



The figure shows the evolution of the condition number for the 10Ind, the 30Ind and the 100FF, as well as for the SP100 and the SP500. The condition number was computed based on the shrunken covariance matrix from Ledoit and Wolff (2004), considering a window size of $\tau = 500$ ($\tau = 120$) observations for the S&P Indices (for the 10Ind, 30Ind and 100FF), and rebalancing the portfolio every month.

Figure 12 shows that the mean, the median, as well as the first and third quartiles of the correlation coefficient, across all constituents, are strictly positive for all data sets. In fact, the correlations of the Industry and Fama French portfolios are high and positive

[7]The condition number of the covariance matrix is defined as the ratio of the largest to the smallest eigenvalue (Meucci 2005).

Figure 12: Correlation Coefficients in Time




The figure plots the mean, the median, the first and the third quantile of the correlation coeffcients across all constituents for the 10Ind, 30Ind, and 100FF, as well as for the SP100 and SP500, considering a windowsize of $\tau = 120$ monthly and $\tau = 500$ daily observations, and re-balancing the portfolio every month over the period from 01/1970 to 01/2017 and 12/2004 to 01/2016, respectively. The correlation coefficients are computed based on the robust covariance matrix estimate by Ledoit and Wolff (2004).

in the late 1970s, and slightly decrease until the beginning of the millenium, probably due to the new economy boom and the offset of the technology sector. However, with the burst of the DotCom Bubble and later with the beginning of the financial crisis in 2007, the correlation rises sharply for all three data sets. As for the Industry and Fama French portfolios, the values for the SP100 and SP500 increase with the onset of the financial crisis in 2007. After falling slightly, the correlation rises again in 2012, during the European sovereign debt crisis. The correlation coefficient plays an important role for our following analysis, as increased positive correlation among the constituents is reported to reduce the effects of diversification (Choueifaty and Coignard 2008, You and Daigler 2010, Giuzio and Paterlini 2016).

For all portfolios, the optimal weights vector, $\hat{\boldsymbol{w}}_t$, depends on the choice of the optimal $\lambda$ parameter value. To select the optimal tuning parameter, we consider a grid of 100 log-spaced values of $\lambda$ between $10^{-7.5}$ and $10^{1}$, from which we choose $\lambda_{RIDGE} = \lambda_{LASSO} = \lambda_1 = \alpha\Phi^{-1}\left(1 - \frac{0.01}{2k}\right)$. The remaining elements $i = 2, ..., k$ of the $\lambda$ sequence for SLOPE are as before, equal to $\lambda_i = \alpha\Phi^{-1}\left(1 - \frac{0.01i}{2k}\right)$.

Among the 100 lambda values, we select the optimal tuning parameter for the various strategies in the following ways: As the RIDGE produces portfolios in between the GMV and the EW, we create a grid of six portfolios between these two points and choose the portfolio right before the EW. For the LASSO and the SLOPE, we choose the portfolio that lies between the GMV and GMV-LO solution and which provides us with approximately 30% of active positions. Note that as we increase the tuning parameter, beyond the GMV-LO, SLOPE would move along the no-short sale area towards the EW solution. Still, we do not consider this interval here, as we solve the optimization again with SLOPE and an added long-only constraint (SLOPE-LO), to explicitly exploit the grouping feature, which predominates in this area. Consequently, we then select

for SLOPE-LO from an interval of six portfolios, in between the GMV-LO and EW solution, the one with the largest number of groups. To guarantee that all our portfolios can also be implemented in practice, all weights that are smaller in absolute value than the threshold of 0.05% are set to zero.

Given the optimal portfolio vector $\hat{\boldsymbol{w}}_t$ at time $t$, we compute the out-of-sample mean and the out-of-sample standard deviation, defined as:

$$\hat{\mu}_p = \frac{1}{t} \sum_{i=1}^{t} \hat{\boldsymbol{w}}_t \boldsymbol{R}_{t+1} \tag{21}$$

$$\hat{\sigma}_p = \sqrt{\frac{1}{t-1} \sum_{i=1}^{t} (\hat{\boldsymbol{w}}_t \boldsymbol{R}_{t+1} - \hat{\mu}_p)^2} \tag{22}$$

from which we construct the out-of-sample Sharpe Ratio (SR) as:

$$\widehat{SR} = \frac{\hat{\mu}_p}{\hat{\sigma}_p} \tag{23}$$

To evaluate whether the $\widehat{SR}$ and $\widehat{\sigma_p^2}$ of any portfolio is statistically different from our SLOPE procedure, we use the tests developed by Ledoit and Wolf (2008) and Ledoit and Wolf (2011), respectively.

As frequent re-balancing of a portfolio is costly, we complement our analysis by computing the turnover of each portfolio, defined as:

$$\widehat{TO} = \frac{1}{t} \sum_{i=1}^{t} \|\hat{\boldsymbol{w}}_{t+1} - \hat{\boldsymbol{w}}_t\|_1 \tag{24}$$

Furthermore, we include the following diversification measures: the Diversification Ratio (DR), the weight (WDiv) and the risk diversification (RDiv) measures. The DR is defined as the ratio of the weighted asset volatility to the overall portfolio volatility:

$$\widehat{\text{DR}} = \frac{\sum_{i=1}^{k} \hat{w}_i \hat{\sigma}_i}{\hat{\sigma}_p} \quad , \tag{25}$$

where $\hat{\sigma}_i$ is the $i$-th asset's estimated volatility $\hat{\sigma}_p$ is the estimated portfolio volatility, for which the investor typically prefers a higher value (Choueifaty and Coignard 2008). Finally, both the WDiv and RDiv measure the concentration of the portfolio in terms of weights and risk (Maillard et al. 2010, Roncalli 2013). The WDiv ranges from $\frac{1}{k}$ for a perfectly concentrated portfolio up to 1 for the equally weighted portfolio. It is computed according to:

$$\widehat{\text{WDiv}} = \frac{1}{k \times \sum_{i=1}^{k} \hat{w}_i^2} \tag{26}$$

On the other hand, we obtain the RDiv by substituting the weights for the risk contribution, defined as $\widehat{RC}_i = \hat{w}_i \times \partial_{w_i} \sigma(\hat{\boldsymbol{w}}_i)$, where $\partial_{w_i} \sigma(\hat{\boldsymbol{w}}_i)$, defines the marginal contribution to risk (MRC) of asset $i$, that is the first derivative of the portfolio variance with respect to portfolio weight $w_i$. The MRC measures the sensitivity of the portfolio variance, given a change in asset $i$-th weight. The RDiv takes a value of 1 for the equally-weighted risk contributions (ERC) portfolio, which is least concentrated in terms of risk contributions

and $\frac{1}{k}$ for a portfolio which is fully concentrated on one asset:

$$\widehat{\text{RDiv}} = \frac{1}{k \times \sum_{i=1}^{k} \widehat{RC}_i^2} \tag{27}$$

Summing up, we prefer values close to one for the WDiv and the RDiv (Cazalet et al. 2014).

## 4.2 Empirical Results

**Industry and Fama French Portfolios.** Table 3 reports the annualized out-of-sample volatility, the annualized out-of-sample SR, the number of active positions, the turnover, and the Value-at-Risk (VaR), evaluated at the 5% significance level, for the 10Ind, 30Ind and the 100FF, using a window size of $\tau = 120$ observations with monthly rebalancing. We indicate portfolios that are statistically different from our SLOPE procedure at the 10%, 5% and 1% level, given the test for the difference in the SR and the volatility, following Ledoit and Wolf (2008) and Ledoit and Wolf (2011).

Looking at the values for the out-of-sample volatility in Table 3, we observe that no portfolio is statistically different from our new SLOPE procedure, across any of the three data sets. Still, SLOPE yields consistently lower variance than any of the EW, ERC, RIDGE or GMV-LO portfolios. Similar performance can be observed for our two portfolio strategies, SLOPE-LO and SLOPE-MV, and when comparing them to the EW, the ERC and the RIDGE.

Simultaneously, the values for the out-of-sample SR, establish SLOPE among the best performing portfolios, with some results being statistically significant. In detail: for

Table 3: Risk- and Return Measures - Industry and Fama French Portfolios

| | Vol. (in %) | | | Sharpe Ratio | | | AP | | | Turnover | | | VaR 5% (in %) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 10Ind | 30Ind | 100FF | 10Ind | 30Ind | 100FF | 10Ind | 30Ind | 100FF | 10Ind | 30Ind | 100FF | 10Ind | 30Ind | 100FF |
| EW | 14.489 | 16.255 | 17.507 | 0.780** | 0.661*** | 0.682*** | 10.000 | 30.000 | 100.000 | 0.000 | 0.000 | 0.000 | −5.880 | −6.418 | −7.182 |
| GMV | 10.911 | 9.153 | 6.028 | 1.116* | 1.319** | 3.309*** | 9.982 | 29.885 | 99.470 | 0.086 | 0.221 | 0.759 | −4.476 | −3.576 | −1.390 |
| GMV-LO | 11.474 | 11.214 | 13.129 | 1.018 | 1.005 | 0.963*** | 5.371 | 8.569 | 9.292 | 0.042 | 0.055 | 0.085 | −4.881 | −4.657 | −5.592 |
| ERC | 13.578 | 15.029 | 16.905 | 0.845 | 0.735*** | 0.718*** | 10.000 | 30.000 | 100.000 | 0.006 | 0.007 | 0.005 | −5.247 | −5.898 | −6.920 |
| RIDGE | 12.288 | 13.504 | 15.304 | 0.945 | 0.840*** | 0.857*** | 9.962 | 29.798 | 97.737 | 0.027 | 0.046 | 0.037 | −4.809 | −5.213 | −6.132 |
| LASSO | 11.423 | 10.876 | 9.664 | 1.021 | 1.043 | 1.765*** | 6.162 | 11.856 | 28.265 | 0.052 | 0.107 | 0.293 | −4.839 | −4.570 | −3.385 |
| SLOPE | 11.429 | 10.939 | 9.961 | 1.024 | 1.045 | 1.665 | 6.303 | 12.551 | 30.155 | 0.049 | 0.099 | 0.254 | −4.864 | −4.541 | −3.744 |
| SLOPE - LO | 11.686 | 11.870 | 13.704 | 0.963* | 0.968 | 0.953*** | 7.299 | 18.465 | 34.616 | 0.104 | 0.206 | 0.403 | −4.826 | −4.753 | −5.639 |
| SLOPE - MV | 12.726 | 12.511 | 14.038 | 0.854** | 0.864** | 0.901*** | 2.609 | 3.389 | 5.002 | 0.018 | 0.022 | 0.056 | −5.282 | −4.955 | −5.381 |

The table reports the out-of-sample Risk and Return Measures for the 10-, 30-, and 100-Portfolios, considering a windowsize of $\tau = 120$ monthly observations and re-balancing the portfolio every month over the period from 01/1970 to 01/2017. Reported are: The annualized out-of-sample volatility, the annualized out-of-sample Sharpe Ratio, the number of active positions (AP), the average total turnover, and the Value at Risk (VaR) evaluated at the 5% significance. Furthermore, we report the significance for the test of the difference in the volatility and the SRs with regard to SLOPE, at the 10%, 5% and 1% level with $^{*}$, $^{**}$ and $^{***}$, respectively.

the Industry portfolios, we observe the highest SR after the GMV, while for the Fama-French portfolios we obtain statistically higher SR than all strategies, except from the GMV and the LASSO. Finally, across all data sets, SLOPE is able to statistically significantly outperform the EW, challenging its widely reported characteristic of a tough benchmark to beat (DeMiguel et al. 2009b). This observation is consistent for SLOPE-LO and SLOPE-MV.

Beside reducing the overall portfolio variance, our goal is to construct sparse portfolios with a low turnover. For that, reconsider that the EW always invests naively in all constituents and thus has the highest possible number of active positions, while by definition reports a turnover value of zero. Similar values are obtained for the ERC, which aims at equalizing the risk contribution of each asset to the overall portfolio risk. These two portfolios are closely followed by the GMV and the RIDGE penalty, whereas the GMV has an exposure to all assets, and estimation errors can enter unhindered into the optimization (see i.e., Ledoit and Wolff (2004)). In fact, looking at Figure 11, the large values of the condition number imply a high instability due to the presence of multicollinearity, leading to extreme changes in the portfolio composition. Consequently, the GMV has the highest turnover values among the non-regularization strategies. The RIDGE, on the other hand, results in more stable asset allocations, despite not setting any asset weight exactly equal to zero. Although both strategies should invest in all assets, the reported number of active positions are slightly reduced, due to our imposed threshold.

Compared to the strategies above, our new SLOPE procedure is able to promote sparse solutions and to reduce the overall portfolio turnover. In fact, we consistently report lower turnover values than the LASSO portfolio, across all three data sets. Especially interesting is the performance of SLOPE-MV: although the investor suffers from an

increase in volatility, she is still able to outperform the ERC, the RIDGE or the EW. Furthermore, SLOPE-MV has the smallest number of active positions, together with the smallest turnover value, among all sparse portfolio methods.

Table 4: Diversification Measures - Industry and Fama French Portfolios

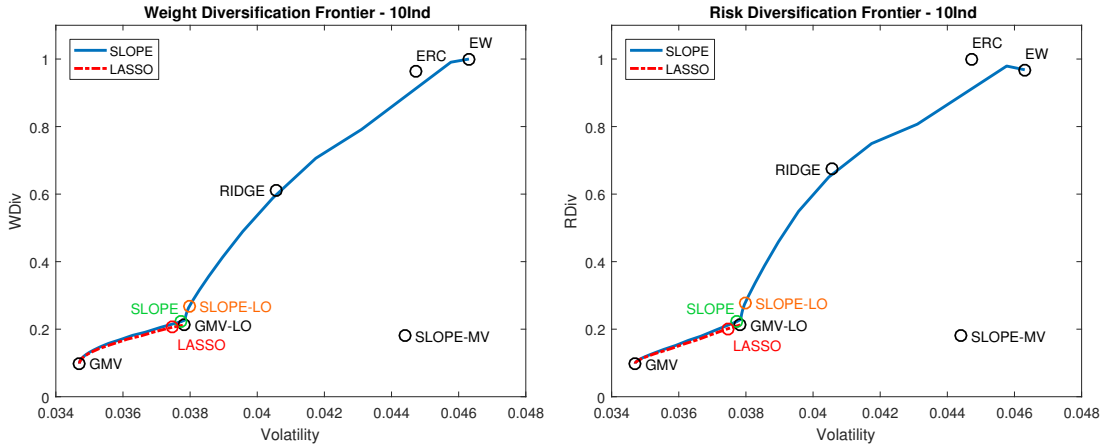
| | DR | | | WDiv | | | RDiv | | |
|---|---|---|---|---|---|---|---|---|---|
| | 10Ind | 30Ind | 100FF | 10Ind | 30Ind | 100FF | 10Ind | 30Ind | 100FF |
| EW | 1.270 | 1.343 | 1.212 | 1.000 | 1.000 | 1.000 | 0.933 | 0.935 | 0.958 |
| GMV | 1.255 | 1.362 | 0.958 | 0.197 | 0.078 | 0.013 | 0.197 | 0.078 | 0.013 |
| GMV-LO | 1.289 | 1.414 | 1.299 | 0.320 | 0.150 | 0.062 | 0.320 | 0.150 | 0.062 |
| ERC | 1.300 | 1.382 | 1.225 | 0.935 | 0.914 | 0.963 | 1.000 | 1.000 | 1.000 |
| RIDGE | 1.328 | 1.430 | 1.244 | 0.655 | 0.629 | 0.622 | 0.735 | 0.722 | 0.617 |
| LASSO | 1.289 | 1.411 | 1.180 | 0.316 | 0.145 | 0.035 | 0.313 | 0.135 | 0.022 |
| SLOPE | 1.293 | 1.424 | 1.200 | 0.328 | 0.157 | 0.043 | 0.326 | 0.148 | 0.027 |
| SLOPE - LO | 1.315 | 1.457 | 1.295 | 0.417 | 0.287 | 0.209 | 0.437 | 0.319 | 0.221 |
| SLOPE - MV | 1.163 | 1.220 | 1.205 | 0.179 | 0.072 | 0.036 | 0.179 | 0.072 | 0.036 |

The table reports the diversification measures for the 10-, 30-, and 100- Portfolios, considering a windowsize of $\tau = 120$ monthly observations and re-balancing the portfolio every month over the period from 01/1970 to 01/2017. Reported are: The Diversification Ratio (DR), the Weight Diversification (WDiv) and the Risk Diversification (RDiv) measures.

In general, our new SLOPE procedure provides the investor with a large amount of flexibility, as with an increased lambda value the penalty starts to form groups among assets, assigning to them the same coefficient value. This is of special interest for investors, who want to move beyond the property of statistical shrinkage, and who want to include in their portfolio construction process any form of financial indicator, like among others fundamental multiples (i.e. Price/Earnings, Dividends/Earnings), accounting values (i.e., net income, Free Cash Flow) or other quantitative measures (i.e., Value-at-Risk or Expected Shortfall). In fact, with SLOPE-MV we here show a simple strategy that selects, out of the formed groups, the one asset with minimum volatility, while other strategies could easily be developed.

Finally, Table 4 complements our risk and return analysis for the 10Ind, 30Ind and 100FF, and reports the DR, the WDiv, and the RDiv. As the EW invests equally in all assets, it achieves, by definition, the best values for the WDiv, with similar values reported for the ERC. As the ERC aims to equalize the contribution to portfolio risk from each asset, it also reports the highest values for the RDiv. SLOPE-LO shows the best diversification measures, followed by SLOPE, while both consistently dominating the LASSO across all data sets. As before, SLOPE does not only outperform the LASSO, but also provides flexibility with regard to the diversification measures. For that, Figure 13 plots the weight- and risk diversification measure against the attainable portfolio volatility for the LASSO and the SLOPE, together with the EW, the ERC, the GMV and the GMV-LO, considering the first window size of $\tau = 120$ observations for the 10Ind.

Figure 13: Risk and Weight Diversification Frontier




The figure shows on the left the weight diversification and on the right the risk diversification frontier, both reporting on the x-axis the portfolio volatility and on the y-axis the risk and weight diversification measure, respectively. Considered are the first window size of $\tau = 120$ months for the 10Ind. Plotted are the resulting combinations for the GMV, the GMV-LO, the EW, the ERC, as well as the different combinations for the LASSO and the SLOPE procedure, considering a range of lambda values from $10^{-7.5}$ to $10^{1}$.

For both frontiers, and considering the solutions of the LASSO, the full grid of lambda parameters enables the investor to select only a combination between the GMV and the

GMV-LO solution. SLOPE, on the other hand, is able to span a much larger set of portfolios, beginning from the GMV, via the GMV-LO up to the EW. The investor can thus control the trade-off between diversification and volatility out of a much larger set of portfolios, to find the allocation that best fits her individual preferences.

**S&P Indices.** Table 5 reports the annualized out-of-sample portfolio volatility, the annualized out-of-sample SR, the number of active positions, the turnover, and the 5% VaR, for the SP100 and SP500, using a window size of $\tau = 500$ daily observations with monthly re-balancing. Again, we indicate portfolios that are statistically different from our SLOPE procedure at the 10%, 5% and 1% level, given the test for the difference in the SR and the volatility, following Ledoit and Wolf (2008) and Ledoit and Wolf (2011).

Table 5: Risk- and Return Measures - S&P Indices

| | Vol. (in %) | | Sharpe Ratio | | AP | | Turnover | | VaR 5% (in %) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | SP100 | SP500 | SP100 | SP500 | SP100 | SP500 | SP100 | SP500 | SP100 | SP500 |
| EW | 18.855 | 20.238 | 0.254 | 0.210 | 93.000 | 443.000 | 0.000 | 0.000 | −7.431 | −8.155 |
| GMV | 10.829 | 11.522 | 0.479 | 0.340 | 92.035 | 434.377 | 0.695 | 2.669 | −5.800 | −6.963 |
| GMV-LO | 12.338 | 10.822 | 0.231 | 0.415 | 18.421 | 30.298 | 0.159 | 0.227 | −6.558 | −4.929 |
| ERC | 16.574 | 17.947 | 0.295 | 0.238 | 93.000 | 443.000 | 0.020 | 0.022 | −6.819 | −7.366 |
| RIDGE | 14.243 | 14.732 | 0.406 | 0.345 | 91.070 | 399.509 | 0.063 | 0.070 | −6.716 | −6.854 |
| LASSO | 11.276 | 9.401 | 0.279 | 0.615 | 36.307 | 146.184 | 0.256 | 0.447 | −6.990 | −4.426 |
| SLOPE | 11.202 | 9.481 | 0.316 | 0.584 | 38.789 | 153.860 | 0.238 | 0.418 | −6.945 | −4.609 |
| SLOPE - LO | 12.474 | 11.991 | 0.392 | 0.402 | 44.991 | 129.632 | 0.464 | 0.584 | −6.197 | −6.221 |
| SLOPE - MV | 13.513 | 13.804 | 0.465 | 0.305 | 10.307 | 13.289 | 0.111 | 0.131 | −5.635 | −5.828 |

The table reports the out-of-sample Risk and Return Measures for the SP100 and SP500, considering a windowsize of $\tau = 500$ daily observations and re-balancing the portfolio every month over the period from 12/2004 to 01/2016. Reported are: The annualized out-of-sample volatility , the annualized out-of-sample Sharpe Ratio, the number of active positions (AP), the average total turnover, and the Value at Risk (VaR) evaluated at the 5% significance. Furthermore, we report the significance for the test of the difference in the volatility and the SRs with regard to SLOPE, at the 10%, 5% and 1% level with $^{*}$, $^{**}$ and $^{***}$, respectively.

Table 5 shows that, with regard to the out-of-sample variance and the SR, no strategy is statistically significantly different from each other. Still, SLOPE and LASSO per-

form best for the SP500, reporting the smallest variance among all strategies and the highest SR. This improved performance can be explained twofold: first, for the SP500 the number of observations is only marginally bigger than the size of our investment universe. Thus, our estimates are very prone to estimation error. We can see that even by using the shrunken covariance matrix, SLOPE and LASSO are still able to reduce extreme weight estimates further. Second, we explicitly select for the LASSO and the SLOPE, a portfolio with a moderate amount of short sales, making it possible to further exploit diversification benefits. Hence, the resulting allocation has a smaller variance, as compared to the GMV-LO. At the same time, and compared to the LASSO, SLOPE is again able to reduce overall turnover, and consequently the cost of implementing such a strategy.

The results for SLOPE-LO and SLOPE-MV are mixed. As we explicitly restrict short sales for these strategies, the resulting variances are higher, but still outperform some of the other allocations. Still, SLOPE-MV reports again the smallest number of active position and the smallest turnover among the two. Depending on the investors objectives, she is then able to further reduce the transaction and monitoring costs.

Finally, Table 6 reports the risk diversification measures for the SP100 and SP500. For the DR the evidence is mixed with SLOPE-LO being superior for the SP100 and SLOPE for the SP500. This difference might be due to short sales in the portfolio, making it possible to exploit more diversification benefits. For the WDiv and RDiv, both SLOPE and SLOPE-LO again dominate the LASSO, hence confirming previous results.

**Robustness Tests.** We investigate the robustness of our results by (a) imposing a linear transaction cost proportional to the turnover and (b) investigating the behavior of the S&P Indices for different window sizes.

Table 6: Diversification Measures - S&P Indices

| | DR | | WDiv | | RDiv | |
|---|---|---|---|---|---|---|
| | SP100 | SP500 | SP100 | SP500 | SP100 | SP500 |
| EW | 1.586 | 1.675 | 1.000 | 1.000 | 0.892 | 0.894 |
| GMV | 1.576 | 3.147 | 0.050 | 0.011 | 0.050 | 0.012 |
| GMV-LO | 1.631 | 1.944 | 0.090 | 0.032 | 0.090 | 0.032 |
| ERC | 1.630 | 1.728 | 0.892 | 0.880 | 1.000 | 1.000 |
| RIDGE | 1.662 | 1.797 | 0.661 | 0.567 | 0.660 | 0.487 |
| LASSO | 1.630 | 2.265 | 0.104 | 0.061 | 0.081 | 0.027 |
| SLOPE | 1.650 | 2.235 | 0.117 | 0.070 | 0.091 | 0.031 |
| SLOPE-LO | 1.706 | 1.936 | 0.285 | 0.206 | 0.312 | 0.219 |
| SLOPE-MV | 1.474 | 1.528 | 0.062 | 0.013 | 0.062 | 0.013 |

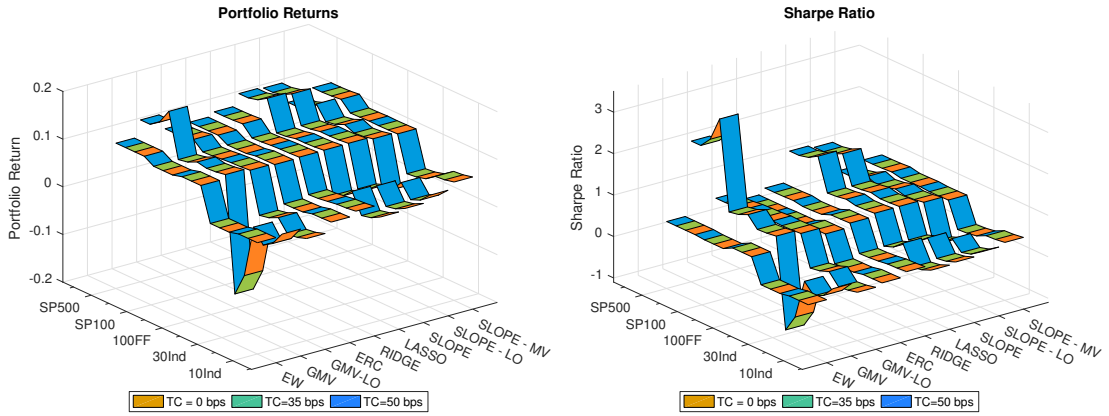
The table reports the diversification measures for the SP100 and SP500, considering a windowsize of $\tau = 500$ daily observations and re-balancing the portfolio every month over the period form 12/2004 to 01/2016. Reported are: The Diversification Ratio (DR), the Weight Diversification (WDiv) and the Risk Diversification (RDiv) measures.

For our transaction cost (TC) analysis, we consider three cost regimes: (1) no (TC = 0bps), (2) low (TC = 35bps) and (3) high (TC = 50bps) transaction costs.[8] Our set-up assumes that the TCs are linear in the turnover and are the same for selling and buying securities. Figure 14 shows the impact of an imposed market transaction cost for the five data sets, and across the different portfolio strategies, as well as considering all three regimes. Naturally, high turnover strategies like the GMV suffer with regard to returns and the SR in the higher cost regimes and even report negative returns for the 30Ind. On the other hand, SLOPE portfolios show a nearly steady performance for all data sets and when considering the different TC regimes.

Finally, we also investigate the performance of the portfolio methods across different window sizes. In detail, we consider a length of $\tau = 250$, $\tau = 750$, and $\tau = 1000$ for the S&P Indices, as well as $\tau = 60$ for the 10Ind, 30Ind and 100FF portfolios. The

[8]Note that, 1 basis point (bps)= 0.01%.

Figure 14: Transaction Cost Regimes




The figure reports the portfolio return and Sharpe Ratio for the different transaction cost regimes across all considered strategies and data sets. We report results for three linear cost regimes, considering (1) no costs (TC = 0bps), (2) low costs (TC = 35bps) and (3) high cost (TC = 50bps), which all are assumed to be linear and the same for buying and selling securities.

results show that our analysis is robust to changes in the window size and due to space limitations, we do not report them here, but make them available from the authors upon request.

# 5 Conclusion

Regularization methods have gained increased attention in the financial literature, because they allow to reduce the influence of estimation errors and to stabilize the resulting asset weight vector. In this paper, we extend the literature on financial regularization by introducing SLOPE to the mean-variance portfolio optimization, discussing its properties and testing its performance with regard to risk and return on simulated and real world data.

SLOPE relies on a sorted $\ell_1$-Norm, whose intensity is controlled by a decreasing sequence of $\boldsymbol{\lambda}$ parameters. As the largest tuning parameter is assigned to the largest

weight, SLOPE penalizes the assets by their rank, and provides a natural interpretation of importance. To solve the penalized mean-variance optimization, we propose a novel algorithm based on the Alternating Direction Method of Multipliers (ADMM). When applied to the LASSO, which is a specific case of SLOPE, this algorithm provides the same accuracy as the state-of-the-art CyCoDe, but is superior with regard to computing time, especially when the asset universe is large.

The paper studies the properties of our new penalty function in a simulated environment and shows that SLOPE has the advantage of still being active in the no short sales area and given an imposed budget constraint. Furthermore, SLOPE can automatically identify assets with the same underlying risk factor exposure and group them together, by assigning the same coefficient value to them. This property is especially desirable for investor planning to incorporate their individual views into the optimization, by selecting assets from these groups according to a specific financial characteristic or individual preferences. We employ such method by introducing a simple investment strategy, SLOPE-MV, selecting from the groups the asset with the lowest volatility.

Moreover, we investigate the performance of SLOPE in an out-of-sample setting, considering a rolling window approach, and re-balancing the portfolio every month. The empirical analysis covers five major data sets, including the 10Ind, 30Ind and 100FF, as well as the constituents of the SP100 and SP500.

Our results show that SLOPE is able to achieve equal and even better out-of-sample portfolio volatilities and SR, when compared to the LASSO. Although, only part of the differences are statistically significant, SLOPE is able to construct sparse portfolios with reduced turnover. This especially applies to situations with a large amount of estimation error, for example when considering the SP500. Furthermore, our SLOPE-MV portfolio results in very sparse portfolios with even lower turnover than state-of-the-art methods

and at the same time maintains a comparable performance.

With regard to weight and risk diversification measures, SLOPE outperforms the LASSO, reporting improved values for the DR, the WDiv and the RDiv. Furthermore, the shape of the penalty extends the frontier of attainable portfolios, ranging from the GMV via the GMV-LO, up to the EW portfolio and thus enables the investor to select among them the one that provides her with the desired volatility- and diversification trade-off.

The results establish SLOPE as a valid alternative to the standard LASSO for creating sparse portfolios with a reduced turnover rate, improved risk- and weight diversification, and a high degree of flexibility in the portfolio construction process.

A natural extension to our study is to investigate, how different sequences of lambda parameters would impact the risk and portfolio allocation, and whether the investor should choose them according to the underlying correlation regime of the stock market or his own prior beliefs on the assets.

# 6 Appendix

## 6.1 ADMM vs. Cyclic Coordinate Descend

In this section, we use the newly introduced ADMM algorithm for solving the minimum-variance optimization with an $\ell_1$ Norm (which is a specific instance of our new SLOPE penalty) and compare its performance to the one of the Cyclic Coordinate Descend algorithm (CyCoDe).

The CyCoDe algorithm is considered state-of-art and has found various applications in solving norm constrained optimization problems (see i.e. Fastrich et al. (2014), Yen (2015)). The algorithm works by optimizing the weights along one coordinate direction, while holding all other weights constant. Although there is no general rule on how the CyCoDe updates the weight vector, we follow the procedure of Yen (2015) and update the weights cyclical, that is we first fix $w_i,\ i = 2, ..., k$ and find a new solution for $w_1$ that is closer to its optimal solution $w^*$. In a next step, we fix $w_i,\ i = 1, 3, ..., k$ and find a value for $w_2$ that is again closer to the optimal one $w^*$. Given a starting criteria $\boldsymbol{w}^0$ for the weight vector, the Lagrange parameter, $\gamma$, for the budget constraint and a trade-off parameter, $\theta$, for $\mu$ and $\sigma^2$, Algorithm 1 shows the pseudo code for the CyCoDe.

**Algorithm 1** Cyclic Coordinate Descend

Initialize $\boldsymbol{w}^0$ and $j = 0$

**while** convergence criteria is not met **do**

  **for** $i = 1$ to $k$ **do**

    $w_i = ST(\gamma - z_i, \lambda) \times (2 \times \sigma_i^2)^{-1}$

    where $ST$ is the soft-thresholding function and $z_i = 2\sum_{j \neq i}^{k} w_j \sigma_{ij} - \theta\mu_i$

  **end for**

  $j = j + 1$

**end while**

To evaluate the performance of the two algorithms, we first draw a random sample of size $n$ for $k$ assets from a multivariate normal $\boldsymbol{X} \sim MVN(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$:

$$\Sigma_{ij} = \begin{cases} 1, & i = j, \\ \rho, & i \neq j, \end{cases} \tag{28}$$

and for which we choose $\rho = 0.2$ and 0.8, respectively. Then, we solve the minimum variance problem given in (2) and subject to the $\ell_1$- Norm on the weight vector, using as an input for $\boldsymbol{\Sigma}$ the shrunken covariance matrix, introduced by Ledoit and Wolff (2004). We initialize both algorithms with a soft starting point $\boldsymbol{w}^0$, that is (1) $\boldsymbol{w}_i^0 = \frac{1}{k}\ \forall\ i = 1, ..., k$, and (2) $\boldsymbol{w}_i^0 = \frac{a_i}{\sum_{i=1}^{k} a_i}$, with $a_i \sim U(0, 1)\ \forall\ i = 1, ..., k$, and repeat the above procedure 100 times, using for both algorithms a tolerance stopping point of $10^{-7}$. All computations are performed in Matlab 2016a on a Lenovo T430, with Windows 7, an Intel i7-3520M with 2.90 GHZ and 8 GB of RAM.

Table 7 and 8 display the minimum and the median of the objective function values, together with the median amount of shorting, the median time in seconds used for each

algorithm to solve the 100 simulations and the median absolute weight difference[9], considering as soft starting criteria an equally weighted and a random portfolio weight vector, respectively.[10]

The tables show that both algorithms reach the same global minimum and median objec-

Table 7: Simulation Results - Equal Weights

| | | | | $\lambda = 4.03 \times 10^{-6}$ | | | | | $\lambda = 5.65 \times 10^{-4}$ | | | | | $\lambda = 7.91 \times 10^{-2}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\rho$ | $n$ | $p$ | Algo | Min | Med | Short | Time | W.Diff. | Min | Med | Short | Time | W.Diff. | Min | Med | Short | Time | W.Diff. |
| 0.2 | 500 | 100 | CyCoDe | 0.14 | 0.16 | 0.51 | 0.66 | $5 \times 10^{-7}$ | 0.14 | 0.16 | 0.49 | 0.62 | $5 \times 10^{-7}$ | 0.23 | 0.25 | 0.00 | 0.18 | $7 \times 10^{-8}$ |
| | | | ADMM | 0.14 | 0.16 | 0.51 | 0.01 | | 0.14 | 0.16 | 0.49 | 0.01 | | 0.23 | 0.25 | 0.00 | 0.01 | |
| | 500 | 250 | CyCoDe | 0.09 | 0.11 | 2.13 | 13.63 | $8 \times 10^{-6}$ | 0.09 | 0.11 | 2.02 | 12.87 | $6 \times 10^{-6}$ | 0.21 | 0.24 | 0.00 | 0.94 | $8 \times 10^{-8}$ |
| | | | ADMM | 0.09 | 0.11 | 2.13 | 0.09 | | 0.09 | 0.11 | 2.02 | 0.09 | | 0.21 | 0.24 | 0.00 | 0.03 | |
| | 1000 | 500 | CyCoDe | 0.09 | 0.10 | 3.46 | 117.69 | $3 \times 10^{-5}$ | 0.09 | 0.11 | 3.23 | 116.29 | $2 \times 10^{-5}$ | 0.22 | 0.24 | 0.00 | 5.58 | $1 \times 10^{-7}$ |
| | | | ADMM | 0.09 | 0.10 | 3.46 | 0.66 | | 0.09 | 0.11 | 3.23 | 0.64 | | 0.22 | 0.24 | 0.00 | 0.17 | |
| 0.8 | 500 | 100 | CyCoDe | 0.55 | 0.64 | 3.39 | 11.67 | $2 \times 10^{-3}$ | 0.55 | 0.65 | 3.30 | 11.23 | $2 \times 10^{-3}$ | 0.73 | 0.83 | 0.00 | 1.37 | $8 \times 10^{-7}$ |
| | | | ADMM | 0.55 | 0.64 | 3.39 | 0.06 | | 0.55 | 0.65 | 3.30 | 0.05 | | 0.73 | 0.83 | 0.00 | 0.03 | |
| | 500 | 250 | CyCoDe | 0.34 | 0.42 | 10.98 | 35.33 | $8 \times 10^{-1}$ | 0.35 | 0.43 | 10.46 | 34.75 | $8 \times 10^{-1}$ | 0.67 | 0.82 | 0.00 | 6.03 | $1 \times 10^{-6}$ |
| | | | ADMM | 0.34 | 0.42 | 10.94 | 0.58 | | 0.35 | 0.43 | 10.47 | 0.56 | | 0.67 | 0.82 | 0.00 | 0.11 | |
| | 1000 | 500 | CyCoDe | 0.36 | 0.42 | 16.49 | 109.37 | 2.1 | 0.38 | 0.44 | 15.44 | 107.64 | 1.8 | 0.75 | 0.83 | 0.00 | 37.20 | $2 \times 10^{-6}$ |
| | | | ADMM | 0.36 | 0.42 | 16.34 | 3.96 | | 0.38 | 0.43 | 15.33 | 3.76 | | 0.75 | 0.83 | 0.00 | 0.61 | |

The table reports, for the Cyclic Coordinate Descend (CyCoDe) and the Alternating Direction Method of Multipliers (ADMM), the simulation results to the penalized minimum variance problem given in (2), considering six data sets drawn from a multivariate normal distribution, with $\rho = 0.2$ and $\rho = 0.8$, respectively, and using the equally weighted portfolio as a soft starting point. Stated are across all 100 simulations: the minimum (Min) and the median (Med) value of the objective function, the median value of the total amount of shorting (Short) the median time in seconds needed to compute the solution (Time) and the average weight difference (W.Diff.).

tive function value and the same amount of shorting for the low correlation environment, regardless of the chosen lambda value and whether we consider the equally weighted or the random weight vector as the soft starting point. This also applies to the low dimensional data set, when the correlation is set to $\rho = 0.8$. When $p = 500$ for $\rho = 0.8$, the ADMM reports a lower amount of shorting for the first two lambda values. This holds regardless of how we choose the soft starting point. This difference might also explain the discrepancy in the weight vectors, which is reported to be the highest for these two data sets. Still, the difference in the resulting weight vectors is modest and amounts to

[9]The difference in the weights is computed as: $\sum |\boldsymbol{w}^{ADMM} - \boldsymbol{w}^{CyCoDe}|$, where $\boldsymbol{w}^{ADMM}$ and $\boldsymbol{w}^{CyCoDe}$ are the optimal weights obtained with the ADMM and the CyCoDe algorithm, respectively.

[10]Due to space limitations, we have restricted ourselves to report the above mentioned measures. Further results, including the standard deviation of the objective function value and the median number of active positions are available upon request to the authors.

an average of $10^{-6}$ for both low correlation environments, and to $10^{-4}$, for the first two high correlation environments and regardless on how we choose the soft starting point. Most notably, the ADMM outperforms the CyCoDe, with regard to the median time in seconds used to compute the solution for all six data sets. This difference is not negligible: the ADMM uses on average 0.265 seconds in the low correlation environment across all lambdas and all starting criteria, while the CyCoDe is slower by a factor of more than 100, using on average 28.88 seconds. This also applies to the high correlation environment, with the ADMM finding the solution, by taking on average 2.65 seconds and the CyCoDe using 38.98 second. Finally, and for both algorithms, selecting the random weight vector as a starting point results in longer computing times, as opposed to using the equally weighted solution.

Figure 15 plots the computing times needed for the CyCoDe and the ADMM for both the

Table 8: Simulation Results - Random Weights

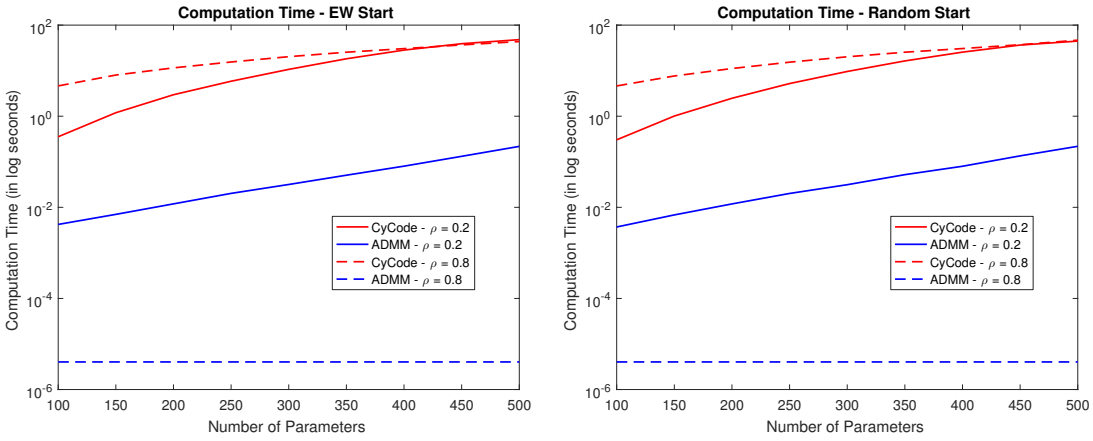
| | | | | $\lambda = 4.03 \times 10^{-6}$ | | | | | $\lambda = 5.65 \times 10^{-4}$ | | | | | $\lambda = 7.91 \times 10^{-2}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\rho$ | $n$ | $p$ | Algo | Min | Med | Short | Time | W.Diff | Min | Med | Short | Time | W.Diff | Min | Med | Short | Time | W.Diff |
| 0.2 | 500 | 100 | CyCoDe | 0.13 | 0.16 | 0.49 | 0.46 | $5 \times 10^{-7}$ | 0.13 | 0.16 | 0.47 | 0.44 | $4 \times 10^{-6}$ | 0.22 | 0.25 | 0.00 | 0.13 | $7 \times 10^{-8}$ |
| | | | ADMM | 0.13 | 0.16 | 0.49 | 0.01 | | 0.13 | 0.16 | 0.47 | 0.01 | | 0.23 | 0.25 | 0.00 | 0.01 | |
| | 500 | 250 | CyCoDe | 0.08 | 0.10 | 2.12 | 10.26 | $8 \times 10^{-6}$ | 0.08 | 0.10 | 2.02 | 10.02 | $6 \times 10^{-6}$ | 0.19 | 0.23 | 0.00 | 0.74 | $8 \times 10^{-8}$ |
| | | | ADMM | 0.08 | 0.10 | 2.11 | 0.07 | | 0.08 | 0.10 | 2.02 | 0.07 | | 0.19 | 0.23 | 0.00 | 0.02 | |
| | 1000 | 500 | CyCoDe | 0.08 | 0.10 | 3.50 | 111.66 | $3 \times 10^{-5}$ | 0.09 | 0.10 | 3.28 | 112.50 | $2 \times 10^{-5}$ | 0.22 | 0.24 | 0.00 | 5.31 | $1 \times 10^{-7}$ |
| | | | ADMM | 0.08 | 0.10 | 3.50 | 0.52 | | 0.09 | 0.10 | 3.28 | 0.51 | | 0.22 | 0.24 | 0.00 | 0.15 | |
| 0.8 | 500 | 100 | CyCoDe | 0.55 | 0.64 | 3.30 | 8.02 | $2 \times 10^{-3}$ | 0.56 | 0.64 | 3.21 | 7.86 | $2 \times 10^{-3}$ | 0.72 | 0.82 | 0.00 | 0.89 | $8 \times 10^{-7}$ |
| | | | ADMM | 0.55 | 0.63 | 3.30 | 0.03 | | 0.55 | 0.64 | 3.21 | 0.03 | | 0.72 | 0.82 | 0.00 | 0.02 | |
| | 500 | 250 | CyCoDe | 0.33 | 0.41 | 10.77 | 31.54 | $8 \times 10^{-1}$ | 0.35 | 0.42 | 10.34 | 32.05 | $8 \times 10^{-1}$ | 0.68 | 0.81 | 0.00 | 5.35 | $1 \times 10^{-6}$ |
| | | | ADMM | 0.33 | 0.41 | 10.75 | 0.55 | | 0.35 | 0.42 | 10.33 | 0.53 | | 0.68 | 0.81 | 0.00 | 0.10 | |
| | 1000 | 500 | CyCoDe | 0.36 | 0.40 | 16.42 | 111.10 | 2.193 | 0.37 | 0.42 | 15.38 | 111.70 | 1.99 | 0.76 | 0.82 | 0.00 | 38.7 | 1.81 |
| | | | ADMM | 0.36 | 0.40 | 16.37 | 3.89 | | 0.37 | 0.42 | 15.36 | 3.69 | | 0.76 | 0.82 | 0.00 | 0.60 | |

The table reports, for the Cyclic Coordinate Descend (CyCoDe) and the Alternating Direction Method of Multipliers (ADMM), the simulation results to the penalized minimum variance problem given in (2), considering six data sets drawn from a multivariate normal distribution, with $\rho = 0.2$ and $\rho = 0.8$, respectively, and using the equally weighted portfolio as a soft starting point. Stated are across all 100 simulations: the minimum (Min) and median (Med) value of the objective function, the median value of the total amount of shorting (Short) the median time in seconds needed to compute the solution (Time) and the average weight difference (W.Diff.).

EW and Random weight vector initialization, considering the two correlation regimes and varying the number of parameters that have to be estimated. Clearly the ADMM consistently shows a superior performance, by only using a fraction of the time of the

CyCoDe. Furthermore, we can observe that both algorithms are also invariant to the selection of the soft starting point. Only the CyCoDe shows a slight difference for parameter values above $k = 450$, signaling that for the CyCoDe an EW portfolio results in finding the optimal solution faster.

Figure 15: Computation Times for CyCoDe and ADMM




The figure shows the average computation times needed for the CyCoDe and ADMM algorithm, depending on the correlation regime, the number of parameters and the soft start criterion. All values are based on 100 simulations, considering a constant correlation set-up.

## 6.2 Portfolio Selection Models

**Equally Weighted Portfolio.** The equally weighted portfolio is considered as one of the toughest benchmarks to beat (see, i.e. DeMiguel et al. (2009b)), and naively distributes the wealth equally among all constituents, such that with $k$ assets:

$$w_i = \frac{1}{k} \; \forall \; i = \{1, ..., k\}, \tag{29}$$

where $w_i$ is the weight of asset $i$. The EW ignores both the variances, the covariances and the return of the assets, and is the optimal portfolio on the mean-variance efficient frontier, when we assume that all three are the same.

**Norm-Constrained Minimum Variance Portfolio**. Reconsider the formulation of the mean-variance problem in (1). By disregarding the mean in the optimization, we obtain the Global Minimum Variance Portfolio (GMV), given by:

$$\min_{\boldsymbol{w}\in\mathbb{R}^k} \sigma_p^2 = \boldsymbol{w}'\Sigma\boldsymbol{w} \quad s.t. \sum_{i=1}^{k} w_i = 1, \ \forall\ i = \{1, ..., k\}, \tag{30}$$

However, this formulation is prone to estimation errors, and unstable portfolio weights. To circumvent these problems, we extend the framework in (30) by adding a penalty function $\rho_\lambda(\boldsymbol{w})$ on the weight vector. For LASSO, we add a $\ell_1$ - Norm to the formulation in (30), such that:

$$\rho_\lambda(\boldsymbol{w}) = \lambda \times \sum_{i=1}^{k} |w_i| \tag{31}$$

where $\lambda$ is a regularization parameter that controls the intensity of the penalty. Besides LASSO, we also consider the RIDGE penalty, which adds an $\ell_2$-Norm on the weight vector to the formulation in (30), and that takes the form of:

$$\rho_\lambda(\boldsymbol{w}) = \lambda \times \sum_{i=1}^{k} w_i^2 \tag{32}$$

As opposed to the LASSO, the RIDGE is not singular at the origin and thus does not promote sparse solutions. Still, imposing the $\ell_2$ - Norm on the portfolio problem is equal to adding an identity matrix, weighted by the regularization parameter $\lambda$ to the

inverse of the variance-covariance matrix, i.e. $(\boldsymbol{\Sigma}^{-1} + \lambda \boldsymbol{I})$, where $\boldsymbol{I}$ is the $k \times k$ identity matrix. This leads to more numerical stability and makes the RIDGE penalty especially appealing in environments that suffer from multicollinearity (Zou and Hastie 2005).

**Equal Risk Contribution Portfolio**. Finally, we consider the Equal Risk Contribution (ERC) portfolio, which aims to equalize the marginal risk contributions of the assets to the overall portfolio risk. That is, given that portfolio variance can be decomposed as:

$$\sigma_p^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} w_i w_j \sigma_{ij} = \sum_{i=1}^{k} w_i \sum_{j=1}^{k} w_j \sigma_{ij} \tag{33}$$

the marginal contribution to the portfolio risk for asset $i$ is given as:

$$c_i^{var} = w_i \sum_{j=1}^{k} w_j \sigma_{ij} = w_i \left(\boldsymbol{\Sigma} \boldsymbol{w}\right)_i \text{ with } \sum_{i=1}^{k} c_i^{var} = \sigma_p^2 \tag{34}$$

where $(\boldsymbol{\Sigma} \boldsymbol{w})_i$ denotes the $i^{th}$ row of the product of $\boldsymbol{\Sigma}$ and $\boldsymbol{w}$ (Roncalli 2013). As the marginal risk is dependent on the portfolio weight magnitude, the ERC portfolio has no analytically solution and must be obtained numerically,by solving:

$$\min_{\boldsymbol{w} \in \mathbb{R}^N} \sum_{i=1}^{k} \left(\frac{w_i \left(\boldsymbol{\Sigma} \boldsymbol{w}\right)_i}{\sigma_p^2} - \frac{1}{k}\right)^2 \; s.t. \; \sum_{i=1}^{k} w_i = 1, \; 0 \leq w_i \leq 1 \; \forall \, i \in \{1, 2, ..., k\} \tag{35}$$

The ERC favors assets with lower volatility, lower correlation with other assets, or both, and is less sensitive to small changes in the covariance matrix as compared to the GMV portfolio (Kremer et al. 2017). Furthermore, (Maillard et al. 2010) show that the volatility of the ERC is between that of the EW and the GMV, and that it coincides with the latter, when both, correlations and SRs, are assumed to be equal (Maillard et al. 2010).

# Acknowledgement

Małgorzata Bogdan acknowledges the grant of the Polish National Center of Science Nr 2016/23/B/ST1/00454, and together with Sandra Paterlini further acknowledge ICT COST Action IC1408 from CRoNoS. Sangkyun Lee acknowledges the support of the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP; Ministry of Science, ICT & Future Planning) (No. 2017R1C1B5018367).

# References

Bellec P, Lecué G, Tysbakov A (2016a) Bounds on the prediction error of penalized least squares estimators with convex penalty. *arXiv:1609.06675, to appear in Modern Problems of Stochastic Analysis and Statistics, Festschrift in honor of Valentin Konakov* .

Bellec P, Lecué G, Tysbakov A (2016b) Slope meets lasso: improved oracle bounds and optimality. *arXiv:1605.08651* 1–29.

Bogdan M, van den Berg E, Sabatti C, Su W, Candes E (2015) Slope - adaptive variable selection via convex optimization. *Annals of Applied Statistics* 9(3):1103–1140.

Bogdan M, van den Berg E, Su W, EJ C (2013) Statistical estimation and testing via the ordered $\ell_1$ norm. *arXiv:1310.1969* 1–46.

Bondell H, Reich B (2008) Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with oscar. *Biometrics* 64(1):115–123.

Boyd S, Parikh N, Chu E, Peleato B, Eckstein J (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning* 3(1):1–122.

Boyle P, Garlappi L, Uppal R, Wang T (2012) Keynes meets Markowitz: The trade-off between familarity and diversification. *Management Science* 58(2):253–272.

Brodie J, Daubechies I, DeMol C, Giannone D, Loris D (2009) Sparse and stable markowitz portfolios. *Proceedings of the National Academy of Science* 106(30):12267–12272.

Carhart M (1997) On persistence in mutual fund performance. *Journal of Finance* 52(1):57–82.

Carrasco M, Noumon N (2012) Optimal portfolio selection using regularization. *Working Paper University of Montreal* 1–52.

Cazalet Z, Grison P, Roncalli T (2014) The smart beta indexing puzzle. *The Journal of Index Investing* 5(1):97–119.

Chen C, Li X, Tolman C, Wang S, Ye Y (2016) Sparse portfolio selection via quasi-norm regularization. *Management Science* 1–34, forthcoming.

Chopra V, Ziemba W (1993) The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management* 19(2):6–11.

Choueifaty Y, Coignard Y (2008) Toward maximum diversification. *Journal of Portfolio Management* 34(4):40–51.

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

DeMiguel V, Garlappi L, Nogales F, Uppal R (2009b) Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies* 22(5):1915–1953.

DeMiguel V, Nogales F (2009) Portfolio selection with robust estimation. *Operations Research* 55(5):798–812.

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

Fan J, Zhang J, You K (2012) Vast portfolio selection with gross-exposure constraint. *Journal of the American Statistical Association* 107(498):592–606.

Fastrich B, Paterlini S, Winker P (2014) Cardinality versus q-norm constraints for index tracking. *Quantitative Finance* 14(11):2019–2032.

Fastrich B, Paterlini S, Winker P (2015) Constructing optimal sparse portfolios using regularization methods. *Computational Management Science* 12(3):417–434.

Fernholtz R, Garvy R, Hannon J (1998) Diversity-weighted indexing. *Journal of Portfolio Management* 4(2):74–82.

Figueiredo M, Nowak R (2014) Sparse estimation with strongly correlated variables using ordered weighted $\ell_1$ regularization. *Working Paper* 1–15, URL `arXiv:1409.4005`.

Gasso G, Rakotomamonjy A, Canu S (2010) Recovering sparse signals with a certain family of non-convex penalties and dc programming. *IEEE Transactions on Signal Processing* 57(12):4686–4698.

Giuzio M, Paterlini S (2016) Un-diversifying during crises: Is it a good idea? *FRB of Cleveland Working Paper* (16-28):1–38.

Hestenes MR (1969) Multiplier and gradient methods. *Journal of Optimization Theory and Applications* 4(5):303–320.

Hoerl A, Kennard R (1988) Ridge regression. Wiley, ed., *Encyclopedia of Statistical Sciences*, volume 8, 129–136 (New York).

Jagannathan R, Ma T (2003) Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance* 58(4):1651–1683.

Kolm P, Tütüncü R, Fabozzi FJ (2014) 60 years of portfolio optimization: Practical challenges and current trends. *European Journal of Operational Research* 234(2):356–371.

Kremer PJ, Talmaciua A, Paterlini S (2017) Risk minimization in multi-factor portfolios: Waht is the best strategy? *Annals of Operations Research* (1-37), URL `http://link.springer.com/article/10.1007/s10479-017-2467-6`.

Ledoit O, Wolf M (2008) Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance* 15(5):850–859.

Ledoit O, Wolf M (2011) Robust performance hypothesis testing with the variance. *Wilmott Technical Paper* 2011(55):86–89.

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

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

Li J (2015) Sparse and stable portfolio selection with parameter uncertainty. *Journal of Business and Economic Statistics* 33(3):381–392.

Maillard S, Roncalli T, Teiletche J (2010) The properties of equally weighted risk contribution portoflios. *The Journal of Portfolio Management* 36(4):60–70.

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

Merton RC (1980) On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8(4):323–361.

Meucci A (2005) *Risk and Asset Allocation* (Springer).

Michaud RO (1989) The markowitz optimization enigma: is 'optimized' optimal? *Financial Analyst Journal* 45(1):31–42.

Parikh N, Boyd S (2014) Proximal algorithms. *Foundations and Trends in Optimization* 1(3):127–239.

Powell MJD (1969) A method for nonlinear constraints in minimization problems. Fletcher R, ed., *Optimization*, 283–298 (New York: Academic Press).

Roncalli T (2013) *Introduction to Risk Parity and Budgeting* (Chapman & Hall/CRC Financial Mathematics Series).

Saab R, Chartrand R, Yilmaz O (2008) Stable sparse approximation via nonconvex optimization.

*33rd International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 3885–3888.

Shefrin H, Statman M (2000) Behavioral portfolio theory. *Journal of Financial and Quantitative Analysis* 35(2):127–151.

Su W, Candès E (2016) Slope is adaptive to unknown sparsity and asymptotically minimax. *Annals of Statistics* 44(3):1038–1068.

Tibshirani R (1996) Regression shrinkage and selection via the lasso. *Royal Statistical Society* 58(1):267–288.

Welsch RE, Zhou X (2007) Applications of robust statistics to asset allocation models. *REVSTAT* 5(1):97–114.

Xing X, Hub J, Yang Y (2014) Robust minimum variance portfolio with $\ell_\infty$ constraints. *Journal of Banking and Finance* 46:107–117.

Yen YM (2015) Sparse weighted norm minimum variance portfolio. *Review of Finance* 20(3):1259–1287.

You L, Daigler R (2010) Is international diversification really beneficial? *Journal of Banking and Finance* 34(1):163–173.

Zeng X, Figueiredo M (2014) Decreasing weighted sorted $l_1$ regularization. *IEEE Signal Processing Letters* 21(10):1240–1244.

Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society* 67(2):301–320.