Quantitative Finance

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rquf20

Sparse index clones via the sorted ℓ_1-Norm

Philipp J. Kremer, Damian Brzyski, Małgorzata Bogdan & Sandra Paterlini

Quantitative Finance, 2022
Vol. 22, No. 2, 349–366, https://doi.org/10.1080/14697688.2021.1962539

Sparse index clones via the sorted ℓ_1-Norm

PHILIPP J. KREMER †, DAMIAN BRZYSKI‡, MAŁGORZATA BOGDAN §¶ and SANDRA PATERLINI *||

†Department of Finance and Accounting, EBS Universität für Wirtschaft und Recht, Wiesbaden, Germany
‡Faculty of Pure and Applied Mathematics, Wroclaw University of Science and Technology, Wroclaw, Poland
§Department of Mathematics, University of Wroclaw, Wroclaw, Poland
¶Department of Statistics, Lund University, Lund, Sweden
||Department of Economics and Management, University of Trento, Trento, Italy

(*Received 10 December 2020; accepted 26 July 2021; published online 15 September 2021*)

Index tracking and hedge fund replication aim at cloning the return time series properties of a given benchmark, by either using only a subset of its original constituents or by a set of risk factors. In this paper, we propose a model that relies on the *Sorted* ℓ_1 *Penalized Estimator*, called SLOPE, for index tracking and hedge fund replication. We show that SLOPE is capable of not only providing sparsity, but also to form groups among assets depending on their partial correlation with the index or the hedge fund return times series. The grouping structure can then be exploited to create individual investment strategies that allow building portfolios with a smaller number of active positions, but still comparable tracking properties. Considering equity index data and hedge fund returns, we discuss the real-world properties of SLOPE based approaches with respect to state-of-the art approaches.

Keywords: Index tracking; Hedge fund clones; Regularization; SLOPE

1. Introduction

Passive replication models are an increasingly popular method to gain exposure to the risk- and return properties of a given benchmark, which could either be a broad market index or an alternative investment vehicle, like a hedge fund.

For replicating equity indices, such models are largely rooted in the empirical evidence that it is not possible to earn a return that is above the average market yield (see i.e. Malkiel 1995, Sorenson *et al.* 1998, Frino and Gallagher 2001). As such, the investor is best advised to just mimic a broad market index, which he could achieve by just buying all of its constituents in the respective relative amounts. However, such a full replication strategy is not only costly, as the manager needs to monitor and frequently rebalance a high number of stocks, but is sometimes not even possible, as small illiquid securities might not be allowed to be traded in large volumes. Furthermore, managerial constraints might prevent the investor from gaining positions in all constituents (Canakgoz and Beasley 2009, Chiam *et al.* 2013).

Hedge funds, on the other hand, have always attracted the attention of investors, as they are outside of most regulatory frameworks, which enables them to create complex and dynamic trading strategies. As a result, these investments show low and even negative correlations with traditional asset classes, like equities or bonds, and are hence of special interest to an investor that seeks risk diversification. Still, investments into hedge funds often demand an initial high endowment or are subject to restrictions, i.e. the fund is not open to new investors (Giamouridis and Paterlini 2010). Consequently, in the case of a broad equity index or a given hedge fund, an alternative approach for the investor is a replicating model that constructs a sparse, i.e. with only a few active positions, and a stable, i.e. with a low turnover, portfolio that best mimics the risk-and return distribution of the given benchmark. In what follows, we will refer to both approaches as the 'Index Tracking (IT)' framework, and to the clone as the 'tracking portfolio'. We explicitly distinguish between the two approaches where necessary.

To construct a tracking portfolio, the literature has mainly focused on minimizing a so called tracking error measure, for example given by the squared deviations between the benchmark returns and its constituents (Rudolf *et al.* 1999). The underlying idea is that the benchmark can be represented by a linear combination of either the equity index constituents or,

*Corresponding author. Email: sandra.paterlini@unitn.it

for the hedge fund replication, by a set of chosen risk factors.† Finding the optimal portfolio then reduces to a simple linear regression framework, often solved by using ordinary least squares (OLS). However, OLS usually leads to allocations that have an exposure to all constituents of the replicating universe, and is thus costly to implement (Giuzio *et al.* 2018).

An alternative way to obtain a sparse coefficient vector is then to impose a so called cardinality constraint, which limits the number of assets in the tracking portfolio. However, including this constraint leads to optimizations, which are NP hard with local optima and discontinuous search spaces. The investor then needs to rely on search heuristics to solve the IT framework, and research in this area has focused on finding new and more efficient algorithms to solve such optimization problems (see e.g. Gilli and Kellezi 2009, Gilli and Winker 2009, Krink *et al.* 2009, Fastrich *et al.* 2014).

In light of dealing with estimation errors, regularization methods, taking the form of a constraint on the norm‡ of the weight vector, have found widespread attention in portfolio selection in the last 10 years, to create sparse and stable allocations (see e.g. Brodie *et al.* 2009, DeMiguel *et al.* 2009, Fastrich *et al.* 2015).§ In the convex case, one of the most prominent penalties is the Least Absolute Shrinkage and Selection Operator (LASSO), introduced to the statistical literature by Tibshirani 1996. The LASSO penalty consists of an ℓ_1-Norm, whose unit ball takes the shape of a cross-polytope, with singularities at the coordinate axes. This makes the penalty a desirable regularization technique, as it achieves in a single step the two tasks of portfolio selection, which are: (1) variable selection, i.e. choosing a subset of assets in which to invest, and (2) parameter estimation, i.e. selecting how much to invest in the chosen assets. With regard to index replication, Brodie *et al.* (2009) state that it can be used to create sparse replicas and at the same time account for transaction costs. Furthermore, Giamouridis and Paterlini (2010) show that by adding the ℓ_1-Norm to the IT model, it is possible to construct sparse and stable hedge fund clones.

Although the LASSO penalty has gained widespread attention, it suffers from various disadvantages, including a reduced recovery of sparse signals when applied to highly dependent data (Fan and Li 2001), selecting at random from equally correlated assets (Bondell and Reich 2008), and being stuck in the no short-sale area, given an imposed budget constraint (i.e. $\sum_{j=1}^{K} w_j = 1$, where w_j represents the jth asset weight, see DeMiguel *et al.* 2009). The latter is of special interest when replicating equity indices, as for such optimizations a long-only constraint (i.e. $w_i \geq 0$, where w_i is the ith asset weight) is typically considered.

To circumvent some of these problems the literature has focused on non-convex penalties like the Logarithmic Penalty (LOG), Smoothly Clipped Absolute Deviation (SCAD) or the ℓ_q -Norm, such as in Fastrich *et al.* (2014) and Giuzio *et al.* (2018).¶ While non-convex penalties are able to produce clones with a smaller number of active positions than LASSO, they suffer from serious numerical issues. It turns out that state-of-the-art interior point or Coordinate Descend (CoDe) methods easily get stuck in local optima and the investor is better advised to resort to heuristic optimization methods (Giuzio 2017). However, none of these methods guarantee convergence to the global minimum.

Given the shortcomings of LASSO and the optimization burdens of non-convex approaches, we aim to extend the literature on IT in the following ways:

First, we introduce the *Sorted* ℓ_1 *Penalized Estimator*, called SLOPE, to the IT framework. SLOPE was recently proposed in statistics by Bogdan *et al.* (2013) and Bogdan *et al.* (2015) and its penalty takes the form of a sorted ℓ_1 -Norm, using a sequence of tuning parameters and penalizing the assets according to their rank magnitude. At the same time, the SLOPE penalty is still convex and singular at the coordinate axes and thus promotes sparse solutions. Furthermore, SLOPE's unit ball takes the form of a polyhedron, which allows for additional reduction of dimension, by assigning the same weights to constituents with similar partial correlations with the time series of benchmark returns. Early investigations of SLOPE mostly focused on the control of the False Discovery Rate for orthogonal and independent predictors (see e.g. Bogdan *et al.* 2013, 2015, Brzyski *et al.* 2018) and the estimation and prediction properties in the context of sparse multiple regression (see e.g. Su and Candès 2016, Bellec *et al.* 2016, 2018), where it was shown that for some specific decreasing sequences of tuning parameters SLOPE is asymptotically minimax in a variety of settings. The clustering properties of SLOPE and its predecessor, the *Octangle Shrinkage and Selection Operator* (OSCAR, Bondell and Reich 2008), are discussed for example in Bondell and Reich (2008) and Figueiredo and Nowak (2014), who show that these methods are able to cluster strongly correlated features. Kremer *et al.* (2020) apply, to our knowledge as one of the first, the SLOPE penalty in a mean-variance framework, providing a simulation study and extensive empirical evidence using equity data. The authors conclude that SLOPE is able to span the entire efficient frontier, while at the same time exploiting grouping structures among assets, to reduce overall turnover and to provide improved risk- and weight diversification measures, relative to other state-of-the-art approaches.

In this article, we provide new theoretical results (see theorems 2.1 and 2.2), which show that the clustering provided by SLOPE is not merely due to the strong correlation between the assets, but it is driven by the similarity of partial correlations of different assets with the respective index or dependent variable. Subsequently, we illustrate these theoretical properties

† Alternative methods for benchmark replication include, e.g. moment matching or payoff distribution approaches that aim at cloning the return distribution of a given index or hedge fund, by first replicating its quantiles, then pricing a payoff function and finally converting it into a portfolio holding strategy. As we focus on the linear case, we do not employ these methods here and refer to Amenc *et al.* (2010) for an overview.

‡ Define the ℓ_q-Norm as: $\|\boldsymbol{w}\|_q = (\sum_{i=1}^{k} |w_i|^{q})^{\frac{1}{q}}$, with $0 < q \leq 1$. If $q = 1$, then $\ell_1 = \sum_{i=1}^{k} |w_i|$ (LASSO), while for $q = 2$ we have $\|\boldsymbol{w}\|_2 = \sum_{i=1}^{k} w_i^2$ (RIDGE). Note that ℓ_q with $0 < q < 1$ is not a norm but a quasi norm.

§ There is an extensive literature dealing with estimation errors in the mean-variance setting, which Brodie *et al.* (2009) show can be rewritten as a regression problem. The interested reader is also referred to the works of Branger *et al.* (2019), Golosnoy *et al.* (2019), or Mainik *et al.* (2015).

¶ Please refer to Appendix 1 for a description of the LOG and SCAD penalties.

in a simulated environment, in which assets belonging to one group exhibit the same partial correlation with the dependent variable, while being uncorrelated to the remaining universe. In such situations, SLOPE is indeed able to identify and assign the assets to their respective groups and is further able to outperform LASSO with regard to Mean Squared Error (MSE). Furthermore, the procedure assigns higher weights to the groups of assets with larger partial correlation to the dependent variable. Second, we consider real-world data and use the grouping feature to construct tracking strategies, based on selecting only the most important groups of approximately equally important constituents. The resulting analysis confirms that our theoretical results hold in real-life. The considered investment strategies are outlined in more detail in section 4.1.

Our empirical analysis considers a rolling window scheme and compares SLOPE to other convex and non-convex penalties, including LASSO, LOG and SCAD. We thereby aim to replicate, on the one hand, the equity indices of the S&P 100 (SP100), S&P 200 (SP200) and S&P 500 (SP500), and on the other, 26 hedge fund indices.

Our empirical results show that by using the grouping properties of the SLOPE penalty, we can create tracking strategies that lead to sparse replicating portfolios with low turnover and improved tracking statistics, which frequently outperform strategies based on the non-convex SCAD and LOG penalties. Specifically, we show that by selecting the relevant groups based on a maximum value of the average partial correlation with the respective index, the resulting allocation often leads to lower tracking error volatility (TEV) and comparable sparsity with regard to current state-of-art regularization techniques.

The paper is structured as follows: section 1 introduces the sparse index tracking model and the theoretical properties of the SLOPE penalty. Section 2 provides insights on the theoretical properties within a simulated environment, while section 3 presents the empirical results on real-world data. Section 4 concludes. All proofs are reported in Appendix 2.

2. Sparse index tracking models

In IT, the investor wants to create a sparse replicating portfolio that best mimics the return time series of a given benchmark. Let $\boldsymbol{Y} = [Y_1, Y_2, \ldots, Y_T]'$ be the $T \times 1$ vector of benchmark returns, and $\boldsymbol{R} = [\boldsymbol{r}^1, \ldots \boldsymbol{r}^K]$ be the $T \times K$ return matrix of benchmark constituents, with columns $\boldsymbol{r}^j = [r_1^j, r_2^j, \ldots r_T^j]'$, representing the $T \times 1$ vector of time series returns for the jth asset. Then, the investor wants to find the optimal $K \times 1$ weight vector $\boldsymbol{w} = [w_1, \ldots w_K]'$ for the replicating portfolio that minimizes the TEV, given by the squared 2-norm ($\|\cdot\|_2^2$) of the difference of the benchmark and the IT portfolio returns. The weight vector $\boldsymbol{w}$ is assumed to be sparse—with not all K weights different from zero. To introduce sparsity in IT, we consider a regression framework and minimize the following penalized tracking error measure, together with the budget constraint:

$$\arg\min_{\boldsymbol{w}\in\mathbb{R}^K} \quad \|\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{w}\|_2^2 + \rho_\lambda(\boldsymbol{w})$$

$$\mathbf{1}'\boldsymbol{w} = \mathbf{1}, \tag{1}$$

where $\rho_\lambda(\boldsymbol{w})$ is a penalty function, whose intensity is controlled by a tuning parameter λ and $\mathbf{1}'$ is a $K \times 1$ unit vector. Following, we shortly outline the LASSO penalty, before introducing our new version of the Sorted ℓ_1-Norm Penalized Estimator (SLOPE). A detailed description of the remaining considered penalties for comparison, can be found in Appendix 1.

2.1. *Least absolute shrinkage and selection operator*

The combination of the ℓ_2 loss function with the ℓ_1-Norm penalty:

$$\rho_\lambda(\boldsymbol{w}) = \lambda \times \|\boldsymbol{w}\|_1 = \lambda \times \sum_{i=1}^{K} |w_i| \tag{2}$$

was firstly used in Santosa and Symes (1986) in the context of signal processing. Then the procedure was reinvented in Chen and Donoho (1994) and Tibshirani (1996) and introduced to the statistical literature as LASSO (Tibshirani 1996).

The ℓ_1 penalty is convex and singular at the origin, and thus promotes sparsity in the portfolio. The penalty has already found widespread attention in index tracking (see e.g. Brodie *et al.* 2009, Giamouridis and Paterlini 2010, Giuzio *et al.* 2018) and portfolio selection (see e.g. Brodie *et al.* 2009, DeMiguel *et al.* 2009, Fan *et al.* 2012, Carrasco and Noumon 2012, Yen and Yen 2014, Fastrich *et al.* 2015, Yen 2015). Due to its computational simplicity it is typically considered a benchmark, against which new regularization methods are tested (Xing *et al.* 2014).

From an economic perspective, DeMiguel *et al.* (2009) show that given the budget constraint, the LASSO regulates the total amount of short-sales in the portfolio, whereas this amount can arbitrarily be distributed across all active components, as opposed to being individually assigned to some weights. The choice, of which assets are finally penalized or sold short, however, depends on the underlying dependence structure (Fastrich *et al.* 2015).

Still, the LASSO suffers from known shortcomings, such as selecting at random from equally correlated assets (Bondell and Reich 2008), of being biased for large coefficient values, especially in the presence of multicollinearity (Fan and Li 2001, Giuzio and Paterlini 2018), and of being ineffective in the no-short-sale area, i.e. $w_i \geq 0$ (DeMiguel *et al.* 2009).

2.2. *Sorted ℓ_1 penalized estimator*

In this paper, we introduce the *Sorted ℓ_1 Penalized Estimator* (SLOPE) to the IT framework. SLOPE is a convex optimization procedure, which, contrary to LASSO, groups correlated assets, instead of selecting from them at random. Moreover, and again contrary to LASSO, SLOPE is active in the no-short-sale area. The SLOPE penalty takes the form of a sorted ℓ_1-Norm such that:

$$\rho_\lambda(\boldsymbol{w}) = \sum_{i=1}^{K} \lambda_i |w|_{(i)} = \lambda_1 |w|_{(1)} + \lambda_2 |w|_{(2)} + \cdots + \lambda_K |w|_{(K)}$$

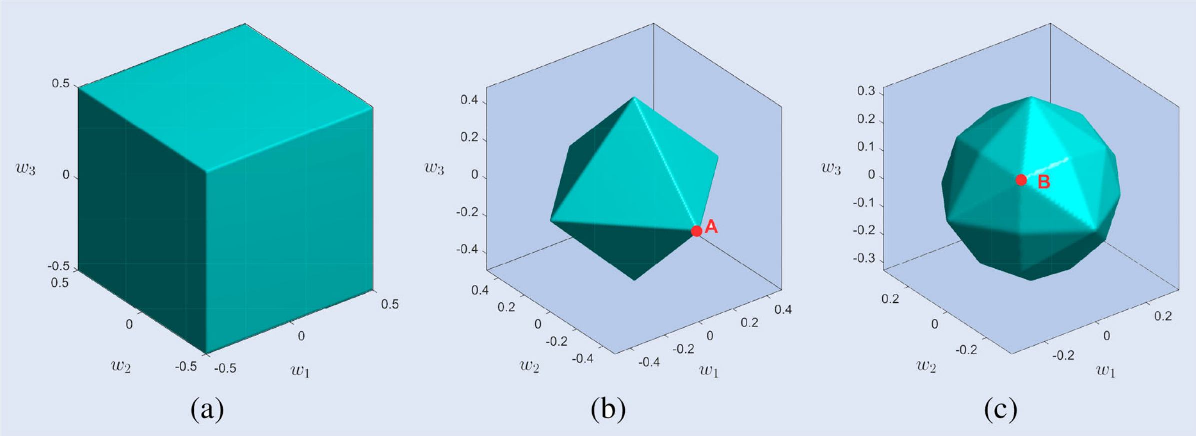


Figure 1. Shapes of SLOPE for a three asset universe.
Notes: The figure plots from left to right, the shapes of unit balls for different SLOPE penalties for a three asset universe, and when considering different setups for the sequence of lambda parameters ($\boldsymbol{\lambda}$). This includes in panel (a) the shape of the ℓ_∞-Norm, with $\boldsymbol{\lambda} = [2\ 0\ 0]$, in panel (b) the shape of the LASSO, with $\boldsymbol{\lambda} = [2\ 2\ 2]$, and in panel (c) the shape of SLOPE with $\boldsymbol{\lambda} = [3\ 2\ 1]$. Point A (Point B) in panel (b) (c) represents a solution in which $w_1 = w_2$, and for which SLOPE groups assets in two groups, depending on the value of w_i. among the assets in the universe.

$$\text{s.t.}\quad \lambda_1 \geq \lambda_2 \geq \cdots \lambda_K \geq\ 0 \text{ and } |w|_{(1)} \geq |w|_{(2)} \geq \cdots |w|_{(K)} \tag{3}$$

where $\boldsymbol{\lambda}_{SLOPE} = [\lambda_1, \lambda_2, \ldots, \lambda_K]$ is a non-increasing sequence of tuning parameters and $|w|_{(i)}$ denotes the ith largest entry of the weight vector $\boldsymbol{w}$ in absolute value. Note that for $\lambda_1 = \lambda_2 = \cdots = \lambda_K$, SLOPE is equivalent to LASSO.

SLOPE requires to choose a sequence of decreasing $\boldsymbol{\lambda}$ parameters, instead of a single one, such as for LASSO, LOG and many other penalties. Depending on the tuning parameter vector, SLOPE penalties can take many different forms.

Panel (a) in figure 1 shows that when we choose $\lambda_1 > \lambda_2 = \lambda_3 = 0$, the unit ball of SLOPE takes the shape of a cube, corresponding to the ℓ_∞-Norm. This norm promotes the grouping of assets, i.e. it encourages solutions when two or more assets have the same coefficient value. For example, let $\boldsymbol{w} = [0.2\ 0.6\ 0.2]'$ be the estimated coefficient vector, then we have three different coefficients, but only two groups, as the value of 0.2 is assigned to two assets. On the other hand, choosing the same value for each lambda parameter, i.e. $\lambda_1 = \lambda_2 = \lambda_3 > 0$, leads to the well known diamond shape of the unit ball of LASSO, given in panel (b) of figure 1. Here point A represents a singular point of the LASSO octahedron, in which the portfolio allocation is sparse, with w_1 and w_3 being equal to zero.

Finally, considering a decreasing sequence of lambda parameters $\lambda_1 > \lambda_2 > \lambda_3 > 0$, leads to the shape of SLOPE, which in panel (c) takes the form of a regular polyhedron in the three dimensional space. Hence, the penalty combines the properties of the ℓ_∞- and ℓ_1-Norm, thereby promoting sparse solutions and grouping of assets. In panel (c), this feature is highlighted by point B, in which $w_1 = w_2$ and $w_3 = 0$. In fact, as the following theorem 2.1 shows, the number of different non-zero coefficients depends on the rank of the return matrix.

Theorem 2.1 *Let $\boldsymbol{R}$ be $T \times K$ matrix of benchmark constituents, $\boldsymbol{Y}$ be T-dimensional vector of benchmark returns and assume that $\hat{\boldsymbol{w}}$ is an unique solution to SLOPE optimization problem,*

$$\arg\min_{\boldsymbol{w}} \frac{1}{2}\left\|\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{w}\right\|_2^2 + \rho_\lambda(\boldsymbol{w}), \tag{4}$$

with a non-increasing sequence $\lambda_1 \geq \cdots \geq \lambda_K \geq 0$. If $\boldsymbol{R}$ is of rank $k \leq K$, then the vector $|\hat{\boldsymbol{w}}|$ contains at most k different non-zero values.

Remark 2.1 In some very rare situations the objective function of SLOPE might not be strictly convex and there might exist infinitely many solutions to the SLOPE optimization problem, with the same value of the objective function. In this case at least one of these solutions satisfies the thesis of theorem 2.1.

Kremer *et al.* (2020) conduct an extensive simulation study, showing that SLOPE is able to identify assets, with the same underlying risk factor exposure, out of a large investment universe, and assigns the same coefficient value to them. This is different to the LASSO, which selects the active weights at random from equally correlated assets. Here, we further formalize and generalize the empirical observations made by Kremer *et al.* (2020). Theorem 2.2 characterizes the SLOPE property of generating groups of assets by revealing its connection with λ sequence and correlations between the columns of R and residuals.

Theorem 2.2 *Let $\hat{\boldsymbol{w}}$ be the SLOPE estimate for $T \times K$ matrix of benchmark constituents, $\boldsymbol{R}$, and T-dimensional vector of benchmark returns, $\boldsymbol{Y}$. Moreover, suppose that all columns of $\boldsymbol{R}$ have been standardized to the same norm, namely it holds $d := \|\boldsymbol{R}_1\| = \cdots = \|\boldsymbol{R}_K\|$ and that the solution satisfies $\hat{\boldsymbol{w}}_1 \geq \cdots \geq \hat{\boldsymbol{w}}_K \geq 0$ (this can always be achieved by permuting columns of $\boldsymbol{R}$ and changing their signs).*

Then, for any $i \in \{1, \ldots, K-1\}$, it holds

$$\hat{\boldsymbol{w}}_i > \hat{\boldsymbol{w}}_{i+1} \implies \boldsymbol{R}_i^\mathsf{T} r_P - \boldsymbol{R}_{i+1}^\mathsf{T} r_P \geq \lambda_i - \lambda_{i+1}, \quad \text{where} \quad r_P := \boldsymbol{Y} - \boldsymbol{R}_{\setminus i,i+1}\hat{\boldsymbol{w}}_{\setminus i,i+1} \tag{5}$$

and $\boldsymbol{R}_{\setminus i,i+1}$ and $\hat{\boldsymbol{w}}_{\setminus i,i+1}$ are obtained by removing ith and $i+1$st columns of $\boldsymbol{R}$ and elements of $\hat{\boldsymbol{w}}$.

The quantity $\boldsymbol{R}_i^\top r_P$ is similar to the classical notion of partial correlation. It measures the correlation between the ith asset and the residual, obtained by subtracting from the vector of benchmark returns, Y, the SLOPE's prediction based on assets other than $\boldsymbol{R}_i$ and $\boldsymbol{R}_{i+1}$. Similarity of $\boldsymbol{R}_i^\top r_P$ and $\boldsymbol{R}_{i+1}^\top r_P$ suggests that these two assets have a similar impact on $\boldsymbol{Y}$ when all other assets are included in the model. Thus, theorem 2.2 says that assets having a similar prediction power with respect to the index are grouped together. This grouping feature of SLOPE will enable the investor to incorporate views into the portfolio selection process, by choosing only the most important groups of assets to track the respective index and assigning the same weights to the assets from the same group. This approach allows to obtain sparse and stable predictive models and will be used in section 2 to create the new index tracking strategy, SLOPE-SLC (see section 4.1).

3. Simulation analysis

Our simulation aims to illustrate the theoretical insights from section 1 by investigating the selection and tracking behavior of SLOPE when the underlying data generating process exhibits a grouping property. Consequently, we analyse, if SLOPE is able to create a tracking portfolio, with the number of groups not larger than the rank of the design matrix (as stated by theorem 2.1), while assigning coefficient values according to theorem 2.2. The latter would result in the same coefficients for assets belonging to the same group and larger coefficient values for those assets, with a larger partial correlation with the dependent variable.

The simulation design follows the set-up from section 1, in which the vector of optimal portfolio weights $\boldsymbol{w}$ is obtained, by solving the regression problem given in (1). Our simulation assumes that the rows of the $T\times K$ return matrix $\boldsymbol{R}$ is obtained by considering an underlying risk factor structure, in which assets belonging to one group have the same exposure to a subset of the factors in the universe. In detail, assume that the universe consists of a total of S risk factors and each asset is represented by a linear combination of those factors. Furthermore, let T be the number of observations, K be the number of assets, and $\boldsymbol{F}_{T\times S}=[\boldsymbol{f}_1\,\boldsymbol{f}_2\,\ldots\,\boldsymbol{f}_S]$, where $\boldsymbol{f}_i$ is the $T\times 1$ vector of returns of the ith risk factor. Moreover, let $\boldsymbol{B}_{S\times K}$ 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. $\boldsymbol{R}$) can be constructed as:

$$\boldsymbol{R}=\boldsymbol{F}\times\boldsymbol{B}+\boldsymbol{\epsilon} \tag{6}$$

where $\boldsymbol{\epsilon}$ is a $T\times K$ matrix of normally distributed error terms. In what follows, we consider the following values for generating the return matrix $\boldsymbol{R}$:

- $T=500$, $K=99$, $S=3$,
- the risk factors $f_1,\ldots,f_S$ are independent from the multivariate standard normal $N(0,I_{S\times S})$ distribution, with $I_{S\times S}$ being an identity matrix,
- the vectors of error terms ϵ_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 $N(0,0.05\times I_{K\times K})$
- the loadings matrix $B_{S\times K}$ is made of exactly 33 copies of each of the following columns: $[0.77\;0.64\;\;0]'$, $[0.9\;\;0\;0.42]'$ and $[0\;\;0.31\;0.64]'$.†
- each column of $\boldsymbol{R}$ is normalized to have norm equal to one.

Finally, given $\boldsymbol{R}$, the data generating process for the index that we aim to replicate is as followed:

$$\boldsymbol{Y}=\boldsymbol{R}\boldsymbol{w}+\nu \tag{7}$$

where ν is a $T\times 1$ matrix of normally distributed error terms, i.e. $\nu\sim\sigma\times N(0,1)$, with $\sigma=0.0015$.

To obtain further insights into the grouping and tracking ability of our SLOPE parameter, we consider two scenarios:

- Scenario 1: $\boldsymbol{w}$ is chosen such that the assets belonging to group 2 have coefficients 2, those belonging to group 3 have coefficients 3, and zero otherwise, i.e.

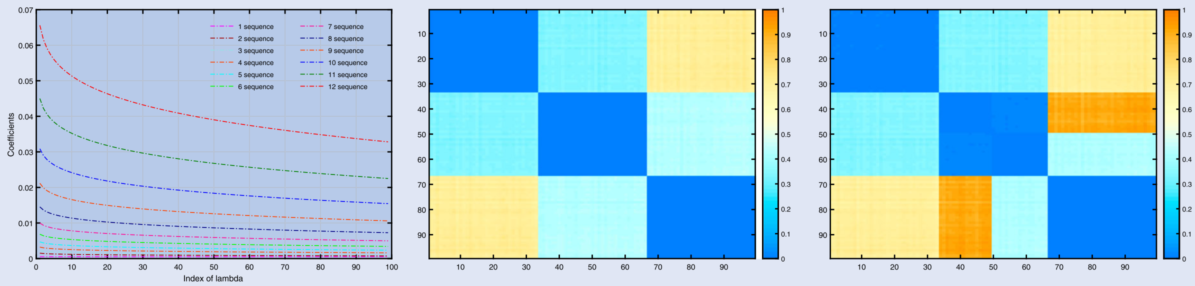
$$\boldsymbol{w}=[\underbrace{0\,0\,\cdots\,0\,0}_{\text{Group1}}\underbrace{2\,2\,\cdots\,2\,2}_{\text{Group2}}\underbrace{3\,3\,\cdots\,3\,3}_{\text{Group3}}]'$$

- Scenario 2: $\boldsymbol{w}$ is chosen, such that all assets in group one have a coefficients of zero, all assets from group 3 have a coefficients of 3, while the first half of assets belonging to group 2 have a coefficients 1 and the other half of assets from group 2 have values equal to 2. That is:

$$\boldsymbol{w}=[\underbrace{0\,0\,\cdots\,0\,0}_{\text{Group1}}\underbrace{1\,1\,\cdots\,1\,1}_{\text{PartI-Group2}}\underbrace{2\,2\,\cdots\,2\,2}_{\text{PartII-Group2}}$$
$$\underbrace{3\,3\,\cdots\,3\,3}_{\text{Group3}}]'$$

To understand, why we have chosen the weight vectors according to scenarios 1 and 2, it is first important to note that the solution to the SLOPE index tracking problem depends on choosing a sequence of tuning parameters that trade-off the minimum tracking error volatility and the number of active weights. For our simulations, we follow Bogdan *et al.* (2013) and choose each component of the sequence of tuning parameters, by setting $\lambda_i=\alpha\Phi^{-1}(1-q_i)$, $\forall i=1,\ldots,k$, where Φ is the standard normal cumulative distribution function and where $q_i=i\times\theta/2k$, with $\theta=0.1$, regulates how fast the sequence of lambda parameters is decreasing. This choice of the sequence of the tuning parameters is motivated by the theoretical results on the asymptotic optimality of the related versions of SLOPE reported in Su and Candès (2016) and Bellec *et al.* (2018) and by the empirical results illustrating their superior prediction properties reported in Bogdan and Frommlet (2021). According to the asymptotic results of Su

† These underlying factor exposures have been chosen specifically to model a block correlation structure among the resulting assets, while ensuring that the determinant of the covariance matrix is non-zero.



(a) Lambda Sequences (b) Scenario 1 (c) Scenario 2

Figure 2. Partial correlation matrix and lambda sequence. (a) Lambda Sequences. (b) Scenario 1 and (c) Scenario 2.
Notes: The figure plots in panel (a) the 12 different sequences of lambda parameters, while panels (b) and (c) show the matrix of absolute differences between SLOPE 'partial correlations' of the respective pairs of assets, i.e. the quantity $M = |\boldsymbol{R}_i^\top r_P - \boldsymbol{R}_{i+1}^\top r_P|$, where$r_P := \boldsymbol{Y} - \boldsymbol{R}_{\setminus i,i+1}\hat{\boldsymbol{w}}_{\setminus i,i+1}$, for lambda sequence number seven, and when we choose $\boldsymbol{w}$ according to scenario 1 and scenario 2, respectively.

and Candès (2016) and Bellec *et al.* (2018), α should be slightly larger than σ, while in Bogdan and Frommlet (2021) it is observed that for real-life data sizes it is usually beneficial to select $\alpha < \sigma$. In our simulation study $\sigma = 0.0015$, therefore we vary α on a grid of 12 log-spaced points between $10^{-3.5} \sim 0.0003$ and $10^{-1.7} \sim 0.02$. Finally, we consider as a starting point, a sequence with no difference in consecutive lambda parameters, i.e. one in which $\lambda_1 = \lambda_2 = \cdots = \lambda_K$, the SLOPE penalty is then equal to the LASSO. Panel (a) of figure 2 shows the resulting 12 different lambda sequences.

Given our chosen set-up and the loading matrix $\boldsymbol{B}_{S\times K}$, we explicitly model a three group block correlation structure for the return matrix $\boldsymbol{R}$, in which assets are allocated in groups, and those belonging to the same group are exposed to exactly two out of the three risk factors. Consequently, assets from the same group are not only highly correlated with each other, but maintaining a low correlation to all other assets. Figure 2(a) shows the resulting block correlation matrix among the assets. Furthermore, given the choice of the weight vector $\boldsymbol{w}$, each group has a specific partial correlation with the index. To illustrate this characteristic and for each of the two scenarios of $\boldsymbol{w}$, figure 2(b,c) plots the quantity M given as:

$$M = |\boldsymbol{R}_i^\top r_P - \boldsymbol{R}_{i+1}^\top r_P|, \quad \text{where } r_P := \boldsymbol{Y} - \boldsymbol{R}_{\setminus i,i+1}\hat{\boldsymbol{w}}_{\setminus i,i+1} \tag{8}$$

As explained in theorem 2.2, the quantity $\boldsymbol{R}_i^\top r_P$ is similar to the classical notion of partial correlation. Consequently, M measures the absolute difference of partial correlations between the assets: the lower the difference, the more similar are the assets in terms of partial correlation with the index. Panels (b) and (c) of figure 2 thus show that, given Lambda Sequence Number 7, we have chosen the values of $\mathbf{w}$ for scenarios 1 and 2 such that we explicitly model a partial correlation block structure with the index, whereas assets belonging to the same group having the same partial correlation with the index and thus a value of M, which is equal to zero.† On the other hand, assets belonging to different groups, exhibit a non-zero value of M. Moreover, in our second scenario we have chosen $\boldsymbol{w}$, in such a way that we further differentiate between the assets from the second group. In fact, half of the assets in group 2 have asset weights equal to 1 and the other half equal to 2. That is, as depicted in figure 2(c), we create a subclass of assets belonging to group 2 with regard to partial correlation and the obtained values of M for those assets. From our insights of theorem 2.2, we thus expect SLOPE to not only group assets with small differences in partial correlation, but also assign higher weights to those assets, for which the partial correlation with the index is higher.

Considering scenario 1, we perform 1000 iterations and report, for each of the 12 lambda sequences, the number of non-zero coefficient and the number of groups, as identified by SLOPE. Furthermore, as a measure of effectiveness, we also compute the Mean Squared Error (MSE) and the Mean Squared Prediction Error (MSPE), given as:

$$MSE(\hat{\boldsymbol{w}}) = \mathbb{E}[\|\boldsymbol{w} - \hat{\boldsymbol{w}}\|_2^2] \tag{9}$$

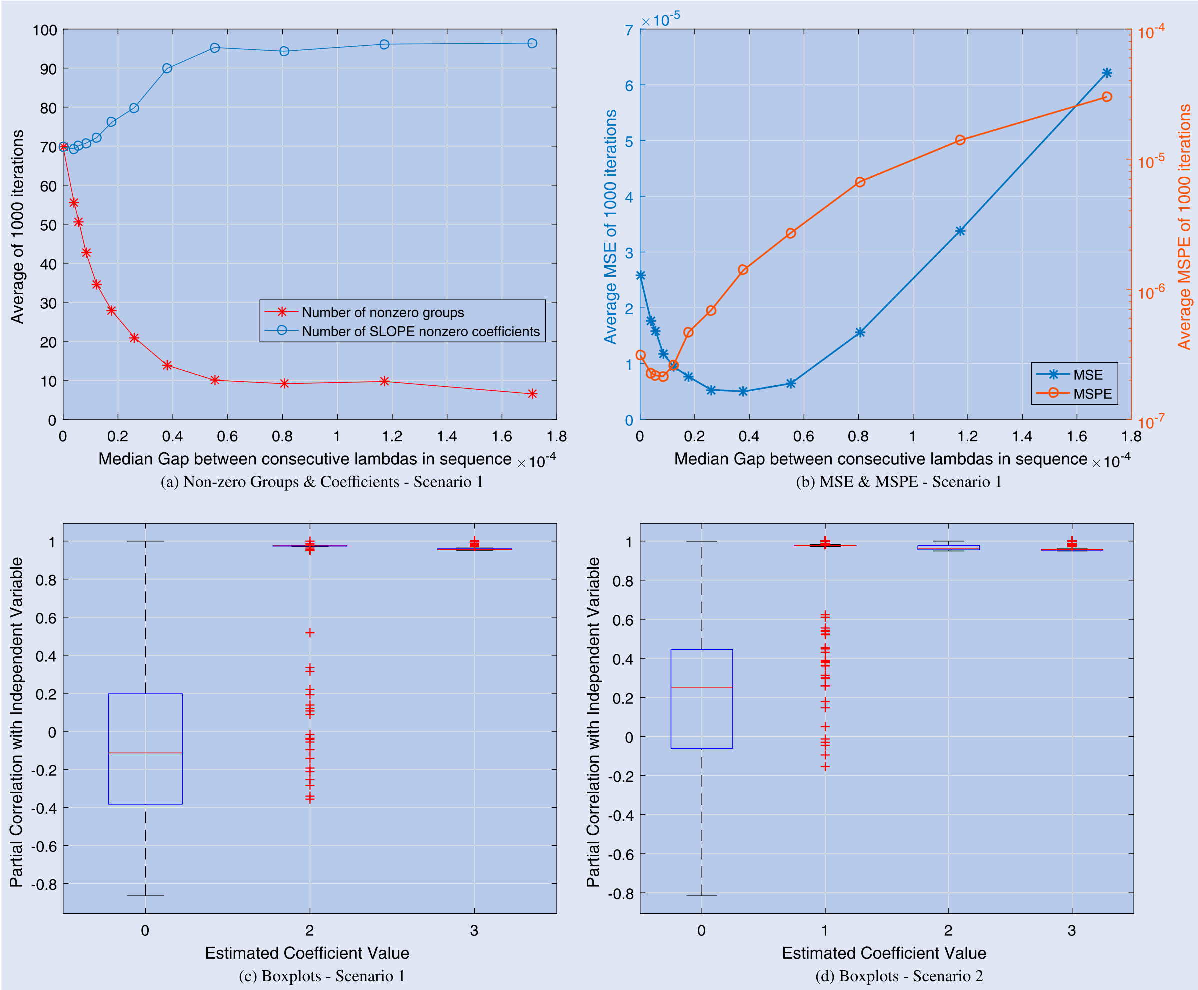
$$MSPE(\hat{\boldsymbol{w}}) = \mathbb{E}[\|\boldsymbol{R}\boldsymbol{w} - \boldsymbol{R}\hat{\boldsymbol{w}}\|_2^2] \tag{10}$$

where $\boldsymbol{w}$ and $\hat{\boldsymbol{w}}$ are equal to the true and estimated portfolio weights, respectively, and $\|\boldsymbol{w} - \hat{\boldsymbol{w}}\|_2^2 = \sum_{i=1}^K (w_i - \hat{w}_i)^2$. Finally, we consider boxplots of group-based partial correlations with the index. That is, over the 1000 iterations, and for a given lambda sequence, we boxplot the partial correlations of those assets which have been assigned the same coefficient value by SLOPE. Here, the SLOPE residual r_P for ith variable is calculated after eliminating all variables from the same cluster.

For the set of lambda sequences (depicted on the x-axis), figure 3 depicts our simulation results, including in panels (a) and (b), the number of non-zero groups, the number of non-zero coefficients, as well as the MSE and the MSPE, respectively. Furthermore panels (c) and (d) show for lambda sequence number seven, i.e. a point that trades of a low MSE and MSPE, the boxplots for scenario 1 and scenario 2, respectively.‡

† These results also hold for the other lambda sequences, and are available from the authors upon request. We here focus on lambda sequence seven, as the higher tuning parameter sequence promotes the grouping of coefficients.

‡ For the sake of brevity, we do not report the results for the Number of non-zero groups, the Number of SLOPE non-zero coefficients, as well as the MSE, and the MSPE for scenario 2 here, but make them available upon request.



(a) Non-zero Groups & Coefficients - Scenario 1

(b) MSE & MSPE - Scenario 1

(c) Boxplots - Scenario 1

(d) Boxplots - Scenario 2

Figure 3. Simulation results. (a) Non-zero groups, (b) MSE and MSPE—scenario 1, (c) boxplots—scenario 1 and (d) boxplots—scenario 2. Notes: The figure plots for the 12 different lambda sequences in panel (a) the number of non-zero coefficients and the non-zero groups, while in panel (b) the MSE and the MSPE. Furthermore, panels (c) and (d) depict, for the lambda sequence number seven, the boxplots of the group based partial coefficients, for those assets which have been assigned the same coefficient value by SLOPE for scenarios 1 and 2, respectively. To create the figures (c) and (d) we round the estimated coefficients of SLOPE to one digit after the comma. All computations are based on 1000 iterations.

From panels (a) and (b) of figure 3, we can observe that as the difference among the consecutive lambda parameters increases, SLOPE starts to form groups among the non-zero coefficients, while the MSE and MSPE are decreasing. Beginning with no difference among the Lambda parameters, a point in which SLOPE is equal to LASSO, solving problem (4) leads to the number of coefficients being equal to the number of groups. In fact, at this point we have no grouping and the procedure results in a randomly sparse weight vector. The latter is a typical behavior of LASSO. As we increase the difference among consecutive lambda values (i.e. we move to the right on the x-axis), SLOPE starts to form groups among the non-zero coefficients. The procedure starts to disentangle the purposely modeled grouping structure, and the MSE and MSPE decrease. As we continue to increase the gaps between consecutive lambda parameters, the octagonal shape of the penalty starts to have a detrimental impact on the estimation, and starts to group all assets together. Consequently, the MSE and MSPE increase. Looking at panels (c) and (d), we boxplot the partial correlation of assets which has been assigned the same coefficient value across all 1000 iterations and considering lambda sequence number seven. As the latter trades-off a low MSE and MSPE, we observe that SLOPE forms groups exactly around the true coefficient values, $\boldsymbol{w}$, considered in scenario 1. Moreover, in scenario 2, SLOPE not only disentangles between the three block type structure that we modeled for $\boldsymbol{R}$, but also disentangles the influence that each of the assets has on $\boldsymbol{Y}$, which confirms the theoretical findings of theorems 2.1 and 2.2.

4. Empirical analysis

4.1. Set up and data

Our empirical analysis investigates the out-of-sample equity index tracking and hedge fund replication ability of the newly introduced SLOPE penalty and compares it to other state-of-art regularization methods, such as LASSO, the SCAD and the LOG penalties. Our ultimate goal is to construct a sparse clone that best tracks the performance of the given benchmark. In addition to the penalty functions, we add the following constraints to the IT model in (1): (a) for tracking the equity indices, the constraint that the asset weights are non-negative (i.e. $w_i \geq 0, \ \forall \ i = 1, \ldots, K$), and (b) for replicating hedge fund returns, which result from complex investment strategies including long and short positions, we restrict the weights to be in the interval $[-1, 1]$. While the equity index replication only considers long-only solutions, we explicitly differentiate between SLOPE and SLOPE-LO for the hedge fund replication, in which SLOPE-LO is SLOPE, together with an added long-only constraint (i.e. $w_i \geq 0$).

Furthermore, in both frameworks we implement a trading strategy, SLOPE-SLC,† in which we utilize the grouping ability of SLOPE, by first solving the IT problem with SLOPE and then keeping only the most important groups of active coefficients. As theorem 2.2 shows, SLOPE groups assets according to their partial correlation with the index. In our search strategy we compute for each group the median partial correlation of the constituents included and keep only those groups active, which have a median partial correlation value above the 75th percentile for the equity indices and the 25th percentile for the hedge fund indices.‡ Then we rescale SLOPE's estimates so that the weights still sum up to 1. The rescaling preserves the group structure, i.e. the assets in the same group still have the same weights. When solving the IT at each t for the hedge fund case, we again distinguish between long-only and short-sales strategies, leading to SLOPE-SLC, and SLOPE-LO-SLC, respectively.§ Finally, and to guarantee that our clones can also be implemented in practice, we impose a threshold and set weights that are smaller in absolute value than 0.05%, to zero.

For our equity index tracking analysis, we focus on reconstructing the daily return observations of the SP100, SP200, and SP500. The data is obtained from Datastream and covers the period from 31 December 2004 to 29 January 2016 ($T = 2890$ daily return observations). Stocks with missing values are dropped from the dataset. Table 1 shows summary statistics for the four equity indices, confirming the typical return time series characteristics of fat tails and light negative asymmetry.

For our hedge fund replication study, we obtain the monthly net of fee returns for 26 Hedge Fund Research (HFR) Indices, over the period from 30 June 1994 to 31 July 2017 ($T = 278$ return observations). The data is obtained from the Hedge Fund Research Database that is considered to be an industry standard in benchmarking the performance of hedge fund strategies.¶ The selected indices follow six broad hedge fund strategy dimensions, including Fund of Fund-, Event Driven-, Equity Hedge-, Emerging Markets-, Total Macro- and Relative Value Strategies, as well as a Fund Weighted Composite. The latter is a collection of over 2000 individual funds, excluding Fund of Funds, serving as an overall benchmark for the hedge fund industry. Here table 2 reports summary statistics for these seven broad hedge fund strategy dimensions. A complete table with summary statistics for all 26 different hedge fund indices is available in the online appendix of the paper.

We can observe that all funds provide a positive return over the considered period, whereas Equity Hedge and Emerging Market Funds are among those strategies with the highest return and risk. The high risk of those strategies is also reflected in the highest minimum monthly return. Macro Total strategies pose another interesting picture, as they show a relatively small standard deviation and also the lowest minimum return. This strategy dimension is also the only one, who has a positive skewness. Finally, all strategies show a leptokurtic distribution.

To replicate the performance of the chosen indices, we select a total of 17 risk factors across five asset classes. The monthly returns for the style factors are obtained from Bloomberg and cover the same period as the hedge fund indices, which is from 30 June 1994 to 31 July 2017. While the selection of the replicating factors in the index tracking problem is naturally the constituents of the index, the selection of factors for hedge fund replication is a crucial step. Our chosen factors represent a subset of those utilized by Giuzio *et al.* (2018), and which ultimately have been selected based on the insights from previous replication studies.

Panel (a) of table 3 provides an overview of the 17 style factors, while panel (b) displays the associated correlation matrix. From the matrix, we observe that factors in the equity asset class are highly correlated, while the remaining constituents have low and even up to negative correlation values. Given that SLOPE is able to group assets with similar partial correlations with the index, we might be able to identify those important assets and to further modify the solution according to any of the strategies introduced above.

To investigate the performance of our regularized replication strategies, we employ a rolling window approach, considering a window size of $\tau = 750$ daily ($\tau = 60$ monthly) observations for the index tracking (hedge fund replication) problem. The rolling window approach works as followed: at time t we use the first τ observations of the benchmark and the replicating constituents to estimate the weights $\boldsymbol{w}_t$, by minimizing the in-sample tracking error subject to the respective penalty and budget constraints. Given our optimal weights, we then compute the out-of-sample excess return between the benchmark and the tracking portfolio as: $(\boldsymbol{Y}_{t+1} - \boldsymbol{R}_{t+1}\boldsymbol{w}_t)$. Finally, we roll the estimation window forward, dropping the last and adding the most recent n observations. The process is then repeated until we reach the end of the data set. For our daily return observations for the S&P Indices, we choose $n = 21$, such that we rebalance the portfolio monthly and obtain a total of $M = 102$ out-of-sample observations. For

† In the following, 'SLC' is used as an acronym for 'select'.

‡ We choose a lower percentile value for the hedge fund indices, as the considered universe is smaller, as compared to the equity indices. For the former, using the 75th percentile would result in a too sparse representations. We make these results available upon request.

§ Alternative investment strategies, exploiting the grouping properties could be set up. Here, we focus on one of the simplest one.

¶ The data and more details on the index construction methodology can be found at www.hfr.com.

Table 1. Descriptive statistics for S&P indices.

Index	k	$\hat{\mu}$ (%)	$\hat{\sigma}$ (%)	$\widehat{\text{med}}$ (%)	$\widehat{\text{min}}$ (%)	$\widehat{\text{max}}$ (%)	$\widehat{\text{skew}}$	$\widehat{\text{kurt}}$
SP100	93	0.023	1.30	0.047	– 9.80	11.60	– 0.240	14.816
SP200	134	0.007	1.10	0.009	– 8.70	5.60	– 0.427	7.839
SP500	443	0.023	1.40	0.046	– 10.70	10.90	– 0.418	13.234

Notes: The table reports descriptive summary statistics for the S&P 100, the S&P 200, and the S&P 500 data set, respectively. Reported are the number of constituents (k), the daily mean ($\hat{\mu}$), the daily standard deviation ($\hat{\sigma}$), the daily median ($\widehat{\text{med}}$), the daily minimum ($\widehat{\text{min}}$), the daily maximum ($\widehat{\text{max}}$), the skewness ($\widehat{\text{skew}}$) and the kurtosis ($\widehat{\text{kurt}}$).

Table 2. Summary statistics for hedge fund strategy dimensions.

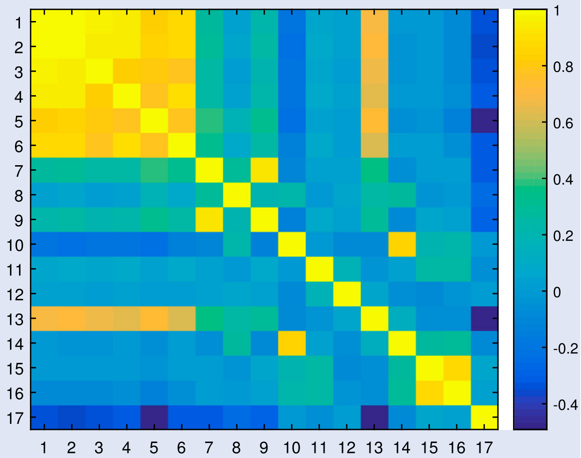
HFRI index	$\hat{\mu}$ (%)	$\hat{\sigma}$ (%)	$\widehat{\text{med}}$ (%)	$\widehat{\text{min}}$ (%)	$\widehat{\text{max}}$ (%)	$\widehat{\text{skew}}$	$\widehat{\text{kurt}}$
Fund weighted composite	0.67	1.92	7.65	– 8.70	0.80	– 0.59	5.99
Fund of funds composite	0.44	1.60	6.85	– 7.47	0.65	– 0.69	7.49
Event driven	0.76	1.87	5.13	– 8.90	1.01	– 1.21	7.16
Equity hedge	0.79	2.54	10.88	– 9.46	0.92	– 0.22	5.31
Emerging markets	0.70	3.78	14.80	– 21.02	1.24	– 0.91	7.61
Macro total	0.59	1.73	6.82	– 3.77	0.42	0.59	3.91
Relative value total	0.63	1.17	3.93	– 8.03	0.76	– 2.75	19.02

Notes: The table reports descriptive summary statistics for the seven hedge fund index strategy dimensions, respectively. Reported are, the monthly mean ($\hat{\mu}$ (%)), the monthly standard deviation ($\hat{\sigma}$ (%)), the monthly median ($\widehat{\text{med}}$ (%)), the monthly minimum ($\widehat{\text{min}}$ (%)), the monthly maximum ($\widehat{\text{max}}$ (%)), the skewness ($\widehat{\text{skew}}$) and the kurtosis ($\widehat{\text{kurt}}$). The values are computed considering the complete time horizon from 30 June 1994 to 31 July 2017.

Table 3. Overview of style factors.

	Equity		Bond
1	S&P 500 Composite	10	Barclays US Treasury
2	Wilshire 5000	11	FTSE World Gov. Bond 1-3Y
3	MSCI USA Value	12	FTSE World Gov. Bond 7-10Y
4	MSCI USA Growth	13	Barclays Global High Yield
5	MSCI EAFE	14	Barclays US MBS
6	MSCI USA Momentum	15	J.P. Morgan US Cash 3M
		16	J.P. Morgan Global Cash 3M
	Commodity		**Currency**
7	S&P GSCI	17	Trade-Weighted US Dollar
8	S&P GSCI Gold		
9	S&P GSCI Crude Oil		

(a)



(b)

Notes: Panel (a) provides an overview of the 17 style factors. Panel (b) shows the correlation between the style factors returns in the period from June 1994 to July 2017. Numbering at the axes corresponds to the numbers in panel (a).

our monthly hedge fund returns, we also rebalance our portfolio monthly and choose $n = 1$ accordingly, leaving us with a total of $M = 218$ out-of-sample observations.†

As for our simulation study, we again need to choose a tuning parameter that trades-off minimum tracking error and the number of active weights, and thus influences the solution to the constrained tracking error problem. In our empirical analysis, we consider for all our penalty functions, a grid of 100 linearly spaced lambda values, while again in the case of SLOPE, these values represent the starting point, $\lambda_{Slope,1}$, of the decreasing sequence of lambda parameters. As before, we follow Bogdan *et al.* (2013) and choose each component of the sequence of tuning parameters, by setting $\lambda_i = \alpha\Phi^{-1}(1 - q_i)$, $\forall i = 1, \ldots, k$, where Φ is the standard normal cumulative distribution function and where $q_i = i \times q/2k$, with $q = 0.1$, regulates how fast the sequence of lambda parameters is decreasing. For our empirical study, we vary the scaling parameter α to consider a grid of starting points $\lambda_1 = \alpha\Phi^{-1}(1 - q_1)$, such that $\lambda_1 = \lambda_{LASSO}$. That is, the shrinkage effect for the first asset is the same as for

† Hedge funds only report their returns on a monthly basis. Furthermore, equity indices only change their composition at the end of a month, at which time stocks are either removed or added to the index. To account for these to effects, and to also ensure a consistent treatment of the two datasets, we choose a monthly rebalancing frequency for both of them.

the LASSO, but is subsequently lower for the $k-1$ securities.

To select the optimal tuning parameter from the set of possible values, the literature has resorted to either using cross-validation techniques or information criteria, like the Akaike Information Criteria (AIC) or the Bayesian Information Criteria (BIC). In this study, we focus on a criterion, inspired by the BIC, to balance the trade-off between the tracking error volatility and the number of active weights and to select the optimal tuning parameters. Using such criteria is preferred to e.g. cross-validation procedures, as it has a lower computational burden (Hastie *et al.* 2001). Therefore, we choose the optimal tuning parameter that minimizes:

$$SC = -2 \times \log\left(\frac{\sum_{t=1}^{\tau}(\boldsymbol{Y}_{t+1} - \boldsymbol{R}_{t+1}\hat{\boldsymbol{w}}_t)^2}{\tau}\right) + \log(\tau) \times \sum_{i}^{k} \mathbb{1}(w_i \neq 0) \tag{11}$$

where $\mathbb{1}$ represents the indicator function and τ is the windowsize. After obtaining our optimal lambda parameter, we evaluate the tracking ability of our clones, by computing the out-of-sample TEV and tracking error (TE) given by:

$$\text{TEV} = \frac{1}{M}\sum_{t=1}^{M}(\boldsymbol{Y}_{t+1} - \boldsymbol{R}_{t+1}\hat{\boldsymbol{w}}_t)^2 \tag{12}$$

$$\text{TE} = \frac{1}{M}\sum_{t=1}^{M}(\boldsymbol{Y}_{t+1} - \boldsymbol{R}_{t+1}\hat{\boldsymbol{w}}_t) \tag{13}$$

where $\boldsymbol{Y}_{t+1}$ is the return of the benchmark and $\boldsymbol{R}_{t+1}$ are the returns of the benchmark constituents at time $t+1$, respectively, while $\hat{\boldsymbol{w}}_t$ is the estimated weight vector obtained from (1). For an ideal tracking portfolio both the tracking error volatility and the tracking error should be close to zero (Giuzio 2017).† Consequently, the optimization problem is set up to minimize the tracking error volatility. Furthermore, we compute the Information Ratio, given by the ratio of the tracking error to the tracking error volatility. As we are interested in a sparse and cost efficient replication of our benchmark, we include statistics on the total number of active positions (AP) and the average total turnover (TO), both given by:

$$\text{AP} = \frac{1}{M}\sum_{t=1}^{M}\sum_{i=1}^{K}\mathbb{1}(w_{i,t} \neq 0) \tag{14}$$

$$\text{TO} = \frac{1}{M}\sum_{t=1}^{M}\sum_{i=1}^{K}|w_{i,t-1} - w_{i,t}| \tag{15}$$

As we want to create tracking portfolios that perform well out-of-sample, we analyse the predictive abilities of our models, by following Gu *et al.* (2020) and compute the predictive R^2_{OOS}, as well as the out-of-sample Maximum Drawdown. The predictive R^2_{OOS} is given as:

$$R^2_{OOS} = 1 - \frac{\sum_{t=1}^{\tau}(\boldsymbol{Y}_{t+1} - \boldsymbol{R}_{t+1}\hat{\boldsymbol{w}}_t)^2}{\sum_{t=1}^{\tau}\boldsymbol{Y}_{t+1}^2}, \tag{16}$$

whereas $-\infty \leq R^2_{OOS} \leq 1$. The out-of-sample Maximum Drawdown is computed as:

$$\text{MaxDD} = \max_{0 \leq t_1 \leq t_2 \leq \tau}(\hat{Y}_{t_1} - \hat{Y}_{t_2}), \tag{17}$$

with $\hat{Y}_{t_i} = (\hat{\boldsymbol{w}}_{t_i-1}R_{t_i})$ and $\hat{Y}_0 = 100$.

Finally, we also test for differences in the out-of-sample predictive accuracy between two models, according to Diebold and Mariano (1995), and calculate the test statistic $DM = \frac{\bar{d}}{\hat{\sigma}_d} \sim N(0,1)$, where $d_{t+1} = (\hat{e}^{(1)}_{t+1})^2 - (\hat{e}^{(2)}_{t+1})^2$, with $\hat{e}^{(m)}_{t+1} = \sum_{t=1}^{\tau}(\boldsymbol{Y}_{t+1} - \boldsymbol{R}_{t+1}\hat{\boldsymbol{w}}^{(m)}_t)$ is the prediction error at time $t+1$ for model m, and $\hat{\boldsymbol{w}}^{(m)}_t$ is the estimated weight vector at time t of that model. Then, $\bar{d}$, and $\hat{\sigma}_d$ denote the mean and the Newey-West standard error of d over the testing period, respectively.

We aim to obtain a portfolio with a low value of tracking error volatility, which is ideally achieved by investing in a small number of active positions. In this context, we do not control for the turnover, whereas this could be improved by using a method similar to Kremer *et al.* (2018), in which the update of the portfolio weights is performed only in cases, in which two consecutive window estimates of the asset's correlation or partial correlation matrices exhibit large variability.‡ In such case we believe the stability of the weight estimates and thereby the turnover could be even further improved.

4.2. *Empirical results*

S&P Indices. Table 4 reports the SP100, SP200 and SP500 indices (by row), reporting, from top left to bottom right, the annualized tracking error volatility, the annualized tracking error, the Information Ratio, the number of active positions, the turnover, the correlation between the replicating portfolio and the given index, as well as the predicted R^2 and the maximum drawdown. Furthermore, we use a t-test to investigate, whether the IR is statistically significantly different from zero at the 1%, 5% and 10% level and report the significance at the IR values. Finally, we report the statistical significance for the Diebold and Mariano (1995) Test at the values for the tracking error.§ All methods are reported in columns that is, from left to right: LASSO, SLOPE-LO, SLOPE-LO-SLC, LOG and SCAD.¶

As described in the previous section, we are interested in constructing a sparse and cost efficient clone that best reflects the performance of the given benchmark. Therefore, the ideal optimal strategy should be based on investing only in a subset of the index constituents (i.e. a small number of active

† The tracking error can be both, positive and negative. A positive (negative) tracking error would indicate that the target index or strategy outperforms (underperforms) the respective tracking clone.

‡ Kremer *et al.* 2018 restrict the rebalancing of the portfolio to instances, in which a Chow-Test indicates that the covariance matrix between rebalancing dates has significantly changed.

§ The test statistic for the IR t-test is computed as: $IR \times \sqrt{M-1}$.

¶ In the table, SLOPE-LO and SLOPE-LO-SLC are displayed as S-LO and S-LO-SLC, respectively.

Table 4. Tracking statistics for S&P indices.

	Tracking error volatility (in %)					Tracking error (in %)					Information ratio					Active positions				
	LASSO	S-LO	S-LO-SLC	LOG	SCAD	LASSO	S-LO	S-LO-SLC	LOG	SCAD	LASSO	S-LO	S-LO-SLC	LOG	SCAD	LASSO	S-LO	S-LO-SLC	LOG	SCAD
SP100	1.80	1.83	3.83	4.03	3.61	1.07	1.29	4.46**	2.18**	2.65***	0.59***	0.702***	1.16***	0.54***	0.73***	83	84	28	48	50
SP200	1.69	1.71	2.12	2.77	3.05	2.40	2.43	2.71	3.76**	3.50***	1.42***	1.42***	1.28***	1.36***	1.15***	103	104	26	25	33
SP500	1.54	1.51	2.74	7.47	14.40	1.76	1.69	3.19***	4.55***	5.95***	1.15***	1.12***	1.16***	0.61***	0.41***	238	262	66	75	74

	Turnover (in %)					Correlation					Predicted R^2_{OOS}					Maximum Drawdown (in %)				
	LASSO	S-LO	S-LO-SLC	LOG	SCAD	LASSO	S-LO	S-LO-SLC	LOG	SCAD	LASSO	S-LO	S-LO-SLC	LOG	SCAD	LASSO	S-LO	S-LO-SLC	LOG	SCAD
SP100	0.08	0.13	0.29	0.29	0.28	0.99	0.99	0.98	0.98	0.98	0.99	0.99	0.93	0.93	0.94	55.03	55.07	59.43	60.59	61.70
SP200	0.07	0.05	0.05	0.14	0.13	1.00	1.00	0.99	0.99	0.99	0.99	0.99	0.98	0.96	0.96	55.29	55.39	53.21	54.73	56.25
SP500	0.04	0.03	0.04	0.08	0.09	1.00	1.00	0.99	0.97	0.94	0.99	0.99	0.97	0.77	0.18	55.55	55.45	56.99	66.97	75.36

Notes: The table reports the out-of-sample tracking statistics for the SP100, SP200, and SP500, respectively. Reported are the annual tracking error volatility (in %)), the annual average percentage tracking error (in %), the Information Ratio, the number of active positions, the average turnover (in %)), the correlation between each respective index and the replicating portfolio, the Maximum Drawdown (in %), as well as the predicted R^2. All values are based on a rolling window analysis, considering a window size of $\tau = 750$ daily observations, between December 2004 and January 2016, and rebalancing the portfolio monthly. Finally, we use ***, **, and * to indicate significance at the 1%, 5% and 10% level, once at the Tracking Error to indicate whether the SLOPE-LO strategy is different to any other strategy according to the Diebold and Mariano (1995) Test, and once at the IR to indicate whether the Information Ratio is significantly different from zero.

positions) with low turnover while having good tracking capabilities with respect to the index, quantified by tracking error volatility, tracking error, correlation with respect to the index, and predicted R^2_{OOS}. From a performance perspective, investment strategies that can track the index and deliver high information ratios and low values of the maximum drawdown would also be attractive.

Focusing on SLOPE-LO, we find that the strategy performs best in terms of tracking error volatility and tracking error for the SP500, while LASSO slightly outperforms for the SP100 and SP200. Nevertheless, both LASSO and SLOPE-LO show the same values of correlation and predicted R^2_{OOS}. In terms of cost efficiency, SLOPE-LO has lower turnover values for SP200 and SP500, achieved with comparable number of active positions for SP100 and SP200, while a larger value for SP500 (i.e. 262 vs. 238). Despite excellent tracking capabilities, both LASSO and SLOPE-LO tend to invest on a much larger number of active positions than SLOPE-LO-SLC, LOG and SCAD. In fact, they invest in about 80% of constituents for SP100 and in about 50% for SP200 and SP500. To improve the cost efficiency of the tracking portfolio, we must either resort to the non-convex penalties, such as SCAD or LOG, or consider the new strategy SLOPE-LO-SLC, which exploits the grouping properties of SLOPE. While SCAD and LOG penalties lead to optimization problems that are NP hard and may have multiple local optima, SLOPE-LO-SLC requires solving a convex problem in a fast and efficient way and then selects only active groups with median partial correlation value above the 75th percentile. In addition to the computational complexity benefits, SLOPE-LO-SLC allows us to identify very sparse portfolios, even sparser than SCAD and LOG for SP100 and SP500, by paying only a small price in terms of the tracking error volatility. In particular, we find that the tracking error volatility of SLOPE-LO-SLC is much smaller for the large problem size in SP500 than for LOG and SCAD (i.e. 2.74 vs. 7.47 and 14.40), while it relies on a smaller number of active positions on average (i.e. 66 vs. 75 and 74). In this case, SLOPE-LO-SLC also has a lower turnover rate (i.e. 0.04), which is even comparable to LASSO (i.e. 0.04) and only slightly larger than SLOPE-LO (i.e. 0.03). When considering the correlation values, SLOPE-LO-SLC performs slightly worse compared to LASSO and SLOPE-LO, but always reaches equal or greater values than LOG and SCAD. In terms of predicted R^2_{OOS}, SLOPE-LO-SLC reports higher values than LOG and SCAD for SP200 and SP500, confirming its properties of providing a sparse and cheap investment strategy, especially for larger problem size, that pays a rather small price in terms of tracking ability.

To give more intuition behind our new SLOPE procedure, figure 4, for SLOPE-LO and for an increasing sequence of lambda parameters, shows in panel (a) the number of groups formed by SLOPE-LO, in panel (b) the number of active positions, in panel (c) the maximum weight, and in panel (d) the group-based median partial correlation with the index. All calculations are based on the first windows of size $\tau = 750$ from daily observations of SP100. In panel (a), one can observe that as the lambda parameter increases, moving from left to right on the x-axis, SLOPE-LO starts to form an increasing number of groups among the assets in the universe. Then, the

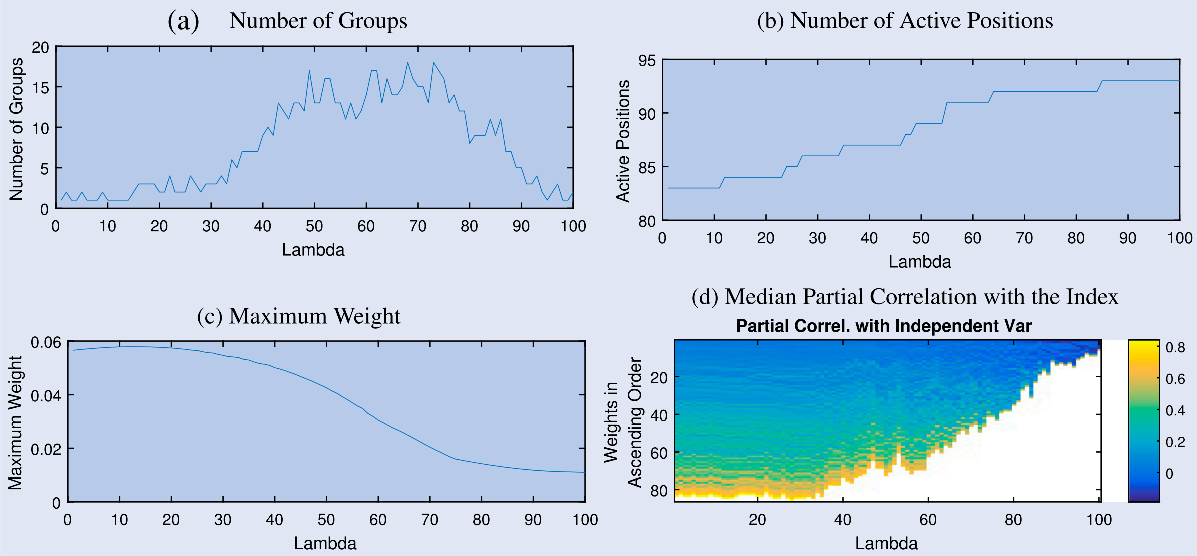


Figure 4. Median partial correlations for SP100. (a) Number of groups, (b) number of active positions, (c) maximum weight and (d) median partial correlation with the index.

Notes: The figure shows for the SP100, using the SLOPE-LO method, considering the first window of size $\tau = 750$ daily observations and varying the lambda parameter between $10^{-4.5}$ and 10^{-2}: in panel (a) the number of groups that SLOPE-LO identifies, in panel (b) the number of active positions that SLOPE-LO identifies, in panel (c) the maximum weight and in panel (d) the median partial correlation across all assets belonging to one group, that is SLOPE-LO has assigned the same coefficient value to them.

number of groups begins to decrease as the octagonal shape of the penalty pushes the solutions toward the equally weighted portfolio, forming larger groups of assets with the same coefficient values until they converge to a single group formed by the equally weighted portfolio for the last value of lambda. Panel (c) shows that for about first 30 values of lambda, the maximum weight values remain approximately constant, while as the value of lambda increases, the maximum weight value begins to decrease toward the weight value of the equally weighted portfolio. Panel (b) consistently shows that the number of active positions continues to grow up to the equally weighted portfolio as lambda increases.† Given the behavior of SLOPE-LO, panel (d) now represents the group-based median partial correlation with the index. That is, for each value of the lambda coefficient, we compute the median partial correlation with the index across all assets assigned the same coefficient value from SLOPE-LO. Our goal is to provide empirical illustration of the theoretical explanations in theorems (1) and (2). To this end, panel (d) not only plots the median partial correlation of all assets belonging to the same group, but also sorts the weights in ascending order on the y-axis . For example, with the smallest lambda value, we obtain 84 distinct non-zero weights. The maximum value is close to 0.06 and the corresponding asset has a partial correlation with the index greater than 0.6. This weight is therefore plotted in yellow near the bottom of panel (d). As the lambda parameter increases and the grouping property of SLOPE-LO becomes stronger, higher coefficient weights initially retain the high partial correlation even in their groups, but then the median partial correlations decrease as the penalty pushes the portfolio toward the equally weighted solutions by assigning the same coefficient value to all assets. Therefore, the investor is best advised to choose a moderate value of lambda and keep only the most important groups from these solutions. SLOPE-LO-SLC is then constructed to retain only the assets within the groups with a median partial correlation above the 75th percentile. This results in a sparser solution with good turnover properties and still reasonable tracking performance compared to LASSO and SLOPE-LO and is able to outperform the non-convex methods in terms of tracking error volatility and tracking error for the large dimensional problems (i.e. SP200, SP500).

Hedge Funds Replication. Hedge fund replication is characterized by portfolios that start from a much smaller number of factors than the number of the index constituents for equity index tracking replication. In fact, in the previous section, optimal portfolio could select among 93, 134 and 443 assets for SP100, SP200 and SP500, respectively. Here, instead we consider 17 potential factors to be used to replicate the returns of 26 different hedge fund indexes. Therefore, sparsity is not then the main relevant aspect in this case, but rather to identify the portfolio that best tracks the hedge fund return and ideally also exhibits some other properties. Table 5 reports the results for the 26 hedge fund indices, including the annualized tracking error Volatility, the annualized tracking error, the Information Ratio, the number of active positions, the turnover, the correlation between the replicating portfolio and the given index, as well as the predicted R^2 and the maximum drawdown. Furthermore, we use a t-test to investigate, whether the IR is statistically significantly different from zero at the 1%, 5% and 10% level and report the significance at the IR values. Similarly, we report the statistical significance for the Diebold and Mariano (1995) Test at the values for the tracking error. As in the equity index tracking analysis, we

† Note that for the solutions of SLOPE-LO the number of active positions increases as we impose a larger tuning parameter. This results from the added long-only constraint, which serves as the initial shrinkage term, and the fact that as lambda increases we move towards the equally weighted portfolio, where all assets are equally weighted.

Table 5. Tracking statistics for hfri hedge fund strategies.

	Tracking error volatility (in %)						Tracking error (in %)						Information ratio						Active positions					
	LASSO	SLOPE	S-SLC	S-LO-SLC	LOG	SCAD	LASSO	SLOPE	S-SLC	S-LO-SLC	LOG	SCAD	LASSO	SLOPE	S-SLC	S-LO-SLC	LOG	SCAD	LASSO	SLOPE	S-SLC	S-LO-SLC	LOG	SCAD
Fund weighted composite	3.17	3.14	3.60	4.42	4.15	3.95	1.98	2.21***	2.61***	1.19***	1.98***	1.74***	0.63***	0.63***	0.32***	0.5***	0.47***	0.58***	8	12	10	16	14	12
Fund of funds composite	3.59	3.54	3.99	5.32	4.78	4.91	0.60*	0.76**	1.27***	− 1.02***	2.32***	− 0.37***	0.15**	0.19***	− 0.05	− 0.09	0.46***	0.25***	8	11	10	15	10	10
FOF conservative	2.85	2.83	3.17	7.71	3.76	4.67	0.30	0.39***	0.08**	− 2.09***	1.99***	− 0.24***	0.12*	0.12*	− 0.08	− 0.17**	0.55***	0.11	7	11	9	12	10	11
FOF diversified	3.83	3.80	4.26	5.48	4.69	5.15	0.54	0.68***	0.69**	− 0.3***	0.91***	0.22***	0.13*	0.16**	− 0.03	− 0.05	0.21***	0.16**	7	11	10	14	10	10
FOF market defensive	5.09	5.14	5.26	6.56	5.44	6.69	1.91	1.40	1.01*	1.02	2.15***	1.17***	0.37***	0.28***	0.37***	0.34***	0.41***	0.23***	6	11	9	12	8	7
FOF strategic	4.64	4.59	4.94	5.09	6.51	5.64	0.87	1.00***	0.22***	0.13***	1.06***	0.94***	0.19***	0.23***	− 0.02	0.05	0.1	0.18**	7	11	10	15	8	8
Event driven	3.47	3.47	3.95	4.37	4.36	4.06	2.05*	2.28***	2.21**	2.11**	1.24*	1.63***	0.6***	0.72***	0.58***	0.61***	0.32***	0.38***	7	11	9	16	9	9
ED distressed restructuring	4.34	4.31	4.52	6.36	5.83	5.66	2.19	2.64**	2.61**	1.58***	3.21***	2.71***	0.52***	0.65***	0.45***	0.56***	0.55***	0.56***	7	10	9	14	6	7
ED merger arbitrage	2.60	2.65	2.89	5.22	3.34	3.59	1.30*	1.26***	1.75**	1.27**	1.51*	1.41***	0.50***	0.48***	0.28***	0.23***	0.48***	0.55***	8	11	10	14	13	11
Equity hedge	4.33	4.25	4.38	4.89	5.22	5.11	1.76	2.01***	2.27***	2.17***	1.99***	1.36***	0.41***	0.5***	0.46***	0.34***	0.35***	0.28***	8	11	10	14	11	10
EH equity market neutral	2.56	2.61	3.80	5.92	3.07	5.23	1.05	0.70**	1.24***	0.04**	0.35*	0.30***	0.39***	0.16**	0.05	0.29***	0.13*	0.46***	7	10	9	14	10	10
EH quantitative directional	5.18	5.11	5.65	7.05	6.71	6.87	3.42	3.25***	3.58***	3.55***	5.30***	3.00***	0.67***	0.61***	0.46***	0.48***	0.72***	0.53***	6	9	8	12	8	7
EH sector technology healthcare	10.07	9.84	9.69	10.44	10.83	10.44	3.42	3.82	3.79	3.60**	3.89	3.08***	0.34***	0.39***	0.38***	0.31***	0.36***	0.34***	6	9	8	10	7	7
Emerging markets	6.64	6.59	6.69	7.15	7.94	7.86	4.63	4.72	4.47**	5.17***	5.55***	4.62***	0.70***	0.72***	0.57***	0.69***	0.69***	0.56***	6	9	8	13	3	3
Emerging markets Asia ex Japan	8.39	8.43	8.40	9.05	9.79	9.12	4.20	4.25	4.06	4.58***	5.07***	3.38***	0.51***	0.52***	0.55***	0.49***	0.50***	0.46***	6	9	7	10	2	3
Emerging markets global	5.71	5.65	5.75	5.92	7.24	6.58	2.73	2.95**	3.20**	3.15***	4.33*	4.35***	0.48***	0.53***	0.42***	0.46***	0.58***	0.71***	6	9	8	13	5	5
Emerging markets Latin America	9.20	9.19	9.18	9.48	9.64	10.27	1.80	1.82	1.67	2.03	− 0.01**	2.13***	0.20***	0.22***	0.20***	0.21***	0	0.15**	6	8	7	9	2	3
Emerging markets Russia Eastern Europe	13.87	13.65	13.70	13.58	14.46	13.84	11.69	11.82	11.78	12.15**	11.64	14.14***	0.84***	0.83***	0.86***	0.86***	0.83***	0.91***	4	6	6	5	2	3
Macro total	5.03	5.08	5.38	6.08	5.48	6.30	1.86	1.50**	1.39	1.29	2.70***	0.44***	0.37***	0.31***	0.31***	0.33***	0.50***	0.25***	7	11	9	12	9	8
Macro systematic diversified	7.60	7.67	7.36	7.97	6.92	8.47	3.08	2.79	2.79	1.54	1.12	0.83***	0.41***	0.44***	0.49***	0.39***	0.17**	0.15**	6	10	9	8	7	7
Relative value total	2.45	2.46	2.70	5.00	3.48	4.35	1.54*	1.66***	1.52**	2.14***	1.77***	1.93***	0.63***	0.80***	0.36***	0.55***	0.59***	0.33***	7	11	9	15	11	10
RV fixed income asset backed	3.56	3.50	3.66	7.10	4.18	5.39	2.8	3.14	3.51	2.53***	2.49***	2.15***	0.79***	0.92***	0.27***	0.72***	0.63***	0.48***	6	9	8	13	8	8
RV fixed income convertible arbitrage	4.31	4.36	4.31	6.03	5.16	5.18	0.34	0.44**	0.52	2.7**	− 1.32	1.87***	0.08	0.09	0.15**	0.15**	− 0.25***	0.26***	6	9	8	16	10	10
RV fixed income corporate	3.41	3.41	3.49	5.42	4.69	4.99	0.57	0.69***	0.17***	− 0.44**	0.25**	− 0.15***	0.16**	0.18**	0.01	0.02	0.03	0.06	6	10	8	13	6	8
RV multi strategy	2.67	2.66	2.76	5.66	4.54	4.55	0.33	0.51**	0.47	0.27***	− 0.36***	− 0.52***	0.12*	0.22***	0.10	0.11	− 0.09	− 0.14*	7	10	9	14	9	9
RV yield alternatives	6.30	6.23	6.17	6.56	7.86	8.27	1.54	1.62	1.58	2.38**	1.70*	2.51***	0.24***	0.22***	0.21***	0.26***	0.19***	0.16**	6	10	9	12	5	5

Fund weighted composite	0.63	1.09	3.41	1.31	1.70	3.27	0.87	0.87	0.82	0.85	0.78	0.80	0.76	0.76	0.41	0.63	0.60	0.62	25.67	25.74	31.98	27.53	26.58	25.73
Fund of funds composite	0.83	1.09	2.36	1.27	2.75	3.77	0.75	0.76	0.71	0.70	0.63	0.65	0.55	0.58	0.28	0.26	0.23	0.29	22.24	23.10	24.79	23.43	25.52	25.10
FOF conservative	0.75	1.18	2.42	1.42	3.98	4.75	0.69	0.69	0.59	0.57	0.52	0.58	0.48	0.48	− 0.12	0.01	0.07	0.18	16.51	17.04	18.97	16.86	19.65	17.13
FOF diversified	0.86	1.09	2.49	1.51	3.15	3.89	0.71	0.72	0.64	0.66	0.61	0.61	0.50	0.51	0.00	0.21	0.25	0.17	22.25	23.36	28.07	23.66	22.68	26.55
FOF market defensive	0.88	1.01	1.80	1.22	3.32	4.51	0.32	0.31	0.27	0.31	0.22	0.19	0.03	0.01	− 0.27	− 0.04	− 0.08	− 0.44	14.30	14.68	15.12	14.98	12.42	17.77
FOF strategic	0.70	1.56	2.23	1.21	1.84	2.76	0.79	0.79	0.76	0.76	0.68	0.71	0.61	0.62	0.46	0.46	0.39	0.43	29.76	29.94	33.46	31.71	28.10	28.10
Event driven	0.67	1.11	2.26	1.22	2.05	4.04	0.84	0.84	0.80	0.81	0.78	0.80	0.72	0.71	0.37	0.48	0.58	0.64	26.49	27.45	37.95	41.56	24.85	25.15
ED distressed restructuring	1.01	1.21	2.45	1.09	2.50	3.63	0.71	0.72	0.67	0.68	0.51	0.57	0.54	0.54	− 0.14	0.25	0.17	0.21	22.76	23.80	47.14	39.05	22.98	21.24
ED merger arbitrage	0.68	1.14	2.00	1.45	3.43	4.12	0.64	0.62	0.57	0.58	0.51	0.61	0.42	0.39	0.09	0.22	0.10	0.30	14.47	14.48	14.30	14.48	17.69	14.50
Equity hedge	0.69	1.44	2.74	1.00	2.41	2.95	0.87	0.88	0.84	0.86	0.82	0.82	0.76	0.77	0.54	0.71	0.65	0.67	33.30	32.87	44.40	35.63	34.25	31.18
EH equity market neutral	0.79	1.14	1.77	1.47	3.74	4.54	0.48	0.46	0.35	0.40	0.18	0.27	0.31	0.28	− 0.93	− 0.33	0.03	− 1.36	7.96	6.60	17.07	16.04	1.87	20.38
EH quantitative directional	0.62	1.03	1.60	0.99	1.95	2.90	0.89	0.89	0.89	0.89	0.86	0.86	0.78	0.79	0.77	0.76	0.66	0.69	46.04	44.13	45.12	47.44	53.73	52.11
EH sector technology healthcare	1.06	1.37	1.76	1.01	3.20	3.90	0.71	0.72	0.71	0.71	0.67	0.70	0.51	0.52	0.50	0.51	0.44	0.49	42.53	42.26	42.26	41.78	44.46	44.01
Emerging markets	0.55	1.01	4.47	0.83	0.77	2.31	0.83	0.84	0.61	0.83	0.78	0.80	0.69	0.70	0.01	0.68	0.56	0.63	44.05	43.71	46.89	45.03	58.18	47.62
EM Asia ex Japan	0.83	0.96	1.60	0.96	1.86	2.67	0.76	0.76	0.76	0.76	0.69	0.72	0.58	0.57	0.57	0.57	0.42	0.49	45.66	45.11	45.49	45.97	56.57	52.32
EM global	0.61	1.00	1.78	1.11	1.07	2.85	0.81	0.82	0.80	0.80	0.76	0.79	0.66	0.66	0.61	0.62	0.45	0.56	35.33	35.28	36.82	37.84	54.81	46.75
EM Latin America	0.71	0.96	1.33	0.87	1.10	2.27	0.77	0.76	0.75	0.76	0.74	0.71	0.59	0.58	0.57	0.58	0.55	0.51	38.94	37.99	40.23	38.94	36.21	32.50
EM Russia Eastern Europe	0.74	1.06	1.14	0.78	0.82	1.15	0.71	0.72	0.72	0.72	0.68	0.72	0.50	0.51	0.51	0.51	0.45	0.49	56.16	55.25	55.25	57.01	61.20	64.35
Macro total	0.84	1.16	2.16	1.14	3.72	4.65	0.45	0.44	0.42	0.44	0.30	0.39	0.16	0.14	0.05	0.12	− 0.01	− 0.02	14.50	14.66	16.22	15.11	8.24	15.07
Macro systematic diversified	0.83	1.89	2.59	1.02	2.77	3.79	0.38	0.38	0.37	0.40	0.45	0.32	0.05	0.03	− 0.05	0.06	0.22	− 0.09	25.33	27.78	31.06	24.77	10.21	19.68
Relative value total	0.68	1.19	2.40	1.27	2.91	3.72	0.80	0.79	0.70	0.78	0.65	0.69	0.69	0.67	0.24	0.62	0.41	0.32	16.09	16.41	16.83	16.27	21.50	14.85
RV fixed income asset backed	0.73	0.80	2.36	1.20	4.69	5.50	0.25	0.25	0.21	0.32	0.08	0.04	0.23	0.24	− 1.70	− 0.24	− 0.02	− 0.33	7.25	6.72	21.75	22.79	4.09	11.75
RV fixed income convertible arbitrage	0.58	1.07	1.67	0.97	1.87	3.13	0.80	0.78	0.75	0.81	0.72	0.73	0.66	0.64	0.59	0.66	0.50	0.56	23.07	21.54	22.32	27.42	18.49	21.65
RV fixed income corporate	0.83	0.95	1.77	0.95	2.99	3.39	0.78	0.79	0.78	0.77	0.62	0.67	0.62	0.64	0.12	0.45	0.31	0.36	22.81	22.80	41.54	30.26	18.46	20.45
RV multi strategy	0.80	0.69	1.99	1.58	3.82	4.34	0.79	0.79	0.70	0.80	0.44	0.61	0.65	0.65	0.28	0.60	0.00	0.19	16.69	16.31	21.33	19.87	14.38	16.09
RV yield alternatives	0.84	1.23	2.12	1.07	2.78	3.37	0.63	0.63	0.64	0.63	0.48	0.50	0.41	0.42	0.39	0.41	0.09	0.07	24.22	22.70	26.14	26.36	30.08	28.77

Notes: The table reports the tracking statistics for the 26 Hedge Fund Indices, covering 6 broad hedge fund strategy dimensions, including Fund of Fund-, Event Driven-, Equity Hedge-, Emerging Markets-, Total Macro- and Relative Value Strategies, as well as a Fund Weighted Composite. Reported are the annual Tracking Error Volatility (in %)), the annual average percentage Tracking Error (in %), the Information Ratio, the number of active positions, the average turnover (in %)), the correlation between each respective index and the replicating portfolio, the Maximum Drawdown (in %), as well as the predicted R^2. All values are based on a rolling window analysis with monthly rebalancing, considering a window size of $\tau = 60$ monthly observations, between June 1994 and July 2017. Finally, we use ***, **, and * to indicate significance at the 1%, 5% and 10% level, once at the TE to indicate whether the SLOPE-LO strategy is different to any other strategy according to the Diebold and Mariano (1995) Test, and once at the IR to indicate whether the Information Ratio is significantly different from zero.

report the indices in the rows and for each computed measure the methods in columns.†

Finally, in the case of hedge fund replication, we not only impose the budget constraint, but also constraint the weights in such a way that they are in the interval $[-1, 1]$. Different to the equity index tracking, we thus explicitly allow short-sales in our replicating portfolios. As outlined by Kremer *et al.* (2020), SLOPE, compared to LASSO, has the desiring property of still being active in the no-short-sale area, where the grouping property of SLOPE is especially dominant and under the presence of a budget constraint. As such, we distinguish, for our newly created tracking strategy, between a short-sale allowed (SLOPE-SLC) and a long-only (SLOPE-LO-SLC) strategy.

Looking at the results from table 5, LASSO and SLOPE reach consistently the lowest out of sample value of tracking error volatility and tracking error among all strategies and considered indices. Given that the available universe of risk factors is small (i.e. 17 in total), as compared to the Equity Index Tracking framework, both strategies achieve this performance additionally with only a small number of active positions. SCAD and LOG penalties do not always converge to sparser solutions than LASSO and SLOPE, while still underperforming in terms of tracking error Volatility. These observations confirm the findings of Giuzio *et al.* (2018), who also found that LOG clones might not always dominate LASSO clones with regard to tracking error volatility, tracking error and turnover. The highest values for the tracking error volatility and tracking error independent of the underlying cloning procedure are reported for the Emerging Market Indices. Those observations can result from changes in the underlying risk exposures and/or structural breaks in managers behavior (see i.e. Fung and Hsieh 2007, Amenc *et al.* 2008). As Emerging Market Funds also make use of macroeconomic changes, this observation is also in line with the findings of Giuzio *et al.* (2018), who show that in an unconstrained regression framework, Global Macro Manager reports the highest turnover, indicating that they frequently change the exposure to different risk factors.

Turning to the 'SLOPE-SLC'—strategies and given the small risk factor universe, the grouping feature does not lead to increased sparsity, as the number of active positions for SLOPE-SLC compared to the LASSO or SLOPE only differ by 1–2 factors. Still, the SLOPE-LO-SLC strategy poses an interesting case. The number of active positions is among the largest across all strategies and indices. Nevertheless, the strategy achieves to reduce the turnover of the overall portfolio and even outperforms the SLOPE and LASSO strategies for some of the Equity Hedge Indices with regard to the tracking error volatility and tracking error. One most notably result are those for the Equity Market Neutral in which the tracking error is only 0.04%, also being significant at the 5% level. The explanation might be that SLOPE in the long-only area pushes the solution to the equally weighted portfolio, thus leading to a higher number of active positions, but a more stable allocation with regard to the gross exposure. Furthermore, given that the risk factor universe consists of a set of six equity risk factors, a broader allocation to those factors might allow for enhanced tracking abilities. Consequently, the fact that SLOPE is still active in the long-only area and pushes the solution towards an allocation with more active positions, might allow the index tracker to capture more return streams and thus allows to improve the tracking ability with a desirable low turnover. As discussed previously, LASSO would be stuck in the long-only case and as it is also evident from the results of table 5 it would then lead to sparser, but not necessarily better hedge fund clones. Notice in fact that the only cases for a negative tracking errors and therefore outperforming the benchmarks are for SLOPE-LO-SLC, which also always exhibit a lower turnover than SLOPE-SLC.

Following the result of table 5, we can observe that SLOPE and LASSO are again able to outperform the more complex, non-convex methods. When considering a limited number of risk factors, as in this investigation, such penalties turn out to be possibly the most appealing. At the same time, SLOPE-SLC procedure is well equipped to perform in line with state-of-the-art tracking portfolios. Considering the idiosyncratic nature of each hedge fund strategy, the grouping feature, and especially the fact that SLOPE is still active in the long-only area might come in handy for index trackers, as it allows to gain a larger exposure to a broader set of underlying risk factors that allow to capture more variation of returns.

5. Conclusion

Index tracking aims at constructing sparse and stable replicating portfolios that best mimics the risk and return time series pattern of a given benchmark, which could either be a broad equity market index or the performance of an alternative investment vehicle, like a hedge fund.

This paper introduces the *Sorted ℓ_1 Penalized Estimator*, called SLOPE, to the index tracking and hedge fund replication framework, and compares its performance to current state-of-the-art convex and non-convex penalty functions. We provide new theoretical insight that SLOPE's grouping ability is based on the difference among the partial correlations of the constituents and that it assigns higher weights to assets which have a larger partial correlation with the respective index. We show these findings in both a simulated and a real-world environment. We find that SLOPE has the desired feature of grouping assets together, enabling us to pick individual constituents from them and hence to create new tracking strategies, such as SLOPE-SLC, that lead to sparse replicating portfolios with good tracking ability, especially for equity index tracking, when the problem typically requires choosing from a large pool of index constituents.

New investment strategies, using the grouping feature can then be developed, as well as testing alternative lambda sequences to improve SLOPE shrinkage and model selection properties. Such extensions are high on our agenda.

† As before, in each table, we denote SLOPE-SLC and SLOPE-LO-SLC as SLOPE-SLC and SLOPE-LO-SLC, respectively.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The research of Malgorzata Bogdan was funded by the NCN [grant number 2016/23/B/ST1/00454]. Research support for Damian Brzyski was provided by the National Institutes of Health [grant number R01MH108467]. Finally, Sandra Paterlini and Malgorzata Bogdan acknowledge support from EU-ICT CRONOS ACTION.

Supplemental data

Supplemental data for this article can be accessed at https://doi.org/10.1080/14697688.2021.1962539.

ORCID

Philipp J. Kremer http://orcid.org/0000-0002-2618-6523
Małgorzata Bogdan http://orcid.org/0000-0002-0657-4342
Sandra Paterlini http://orcid.org/0000-0003-4269-4496

References

Amenc, N., Géhin, W.M. and Meyfredi, J.-C., Passive hedge fund replication: A critical assessment of existing techniques. *J. Altern. Invest.*, 2008, **11**(2), 69–83.

Amenc, N., Martellini, L., Meyfredi, J.-C. and Ziemann, V., Passive hedge fund replication – Beyond the linear case. *Eur. Financ. Manage.*, March 2010, **16**(2), 191–210.

Bellec, P.C., Lecué, G. and Tysbakov, A.B., Bounds on the prediction error of penalized least squares estimators with convex penalty. In *Modern Problems of Stochastic Analysis and Statistics, Festschrift in honor of Valentin Konakov*, edited by V. Panov, 2016.

Bellec, P.C., Lecué, G. and Tsybakov, A.B., Slope meets lasso: Improved oracle bounds and optimality. *Ann. Stat.*, December 2018, **46**(6B), 3603–3642. https://doi.org/10.1214/17-AOS1670

Bogdan, M. and Frommlet, F., Identifying important predictors in large data bases – Multiple testing and model selection. In *Handbook of Multiple Comparisons*, edited by X. Cui, T. Dickhaus, Y. Ding, and J.C. Hsu, 2021 (Chapman & Hall/CRC). Available online at: https://arxiv.org/abs/2011.12154.

Bogdan, M., van den Berg, E., Su, W. and Candès, E.J., Statistical estimation and testing via the ordered ℓ_1 norm. arXiv:1310.1969, pp. 1–46, 2013.

Bogdan, M., van den Berg, E., Sabatti, C., Su, W. and Candes, E.J., SLOPE – Adaptive variable selection via convex optimization. *Ann. Appl. Statist.*, 2015, **9**(3), 1103–1140.

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

Branger, N., Lucivjanska, K. and Weissensteiner, A., Optimal granularity for portfolio choice. *J. Empir. Finance*, 2019, **50**(C), 125–146.

Brodie, J., Daubechies, I., DeMol, C., Giannone, D. and Loris, D., Sparse and stable Markowitz portfolios. *Proc. Nat. Acad. Sci.*, 2009, **106**(30), 12267–12272.

Brzyski, D., Gossmann, A., Su, W. and Bogdan, M., Group slope – Adaptive selection of groups of predictors. *J. Am. Stat. Assoc.*, 2018, **114**(525), 419–433. https://doi.org/10.1080/01621459.2017.1411269

Canakgoz, N.A. and Beasley, J.E., Mixed-integer programming approaches for index tracking and enhanced indexation. *Eur. J. Oper. Res.*, 2009, **196**(1), 384–399.

Candes, E., Waking, M.B. and Boyd, S.P., Enhencing sparsity by reweighted ℓ_1 mminimization. *J. Fourier Anal. Appl.*, 2008, **14**(5), 877–905.

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

Chen, S. and Donoho, D., Basis pursuit. In *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*, Vol. 1, pp. 41–44, 1994 (IEEE).

Chen, C., Li, X., Tolman, C., Wang, S. and Ye, Y., Sparse portfolio selection via quasi-norm regularization. Papers 1312.6350, arXiv.org, 2013.

Chiam, S.C., Tan, K.C. and Al Mamun, A., Dynamic index tracking via multi-objective evolutionary algorithm. *Appl. Soft. Comput.*, 2013, **13**(7), 3392–3408.

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

Diebold, F.X. and Mariano, R.S., Comparing predictive accuracy. *J. Bus. Econ. Stat.*, 1995, **13**, 134–144.

Fan, J. and Li, R., Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.*, 2001, **96**(456), 1348–1360.

Fan, J., Zhang, J. and You, K., Vast portfolio selection with gross-exposure constraint. *J. Am. Stat. Assoc.*, 2012, **107**(498), 592–606.

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

Fastrich, B., Paterlini, S. and Winker, P., Constructing optimal sparse portfolios using regularization methods. *Comput. Manage. Sci.*, 2015, **12**(3), 417–434.

Fernholtz, R., Garvy, R. and Hannon, J., Diversity-weighted indexing. *J. Portfolio Manage.*, 1998, **4**(2), 74–82.

Figueiredo, M. and Nowak, R., Sparse estimation with strongly correlated variables using ordered weighted ℓ_1 regularization. arXiv:1409.4005, Working Paper, pp. 1–15, 2014.

Frino, A. and Gallagher, D., Tracking S&P Index funds. *J. Portfolio Manage.*, 2001, **28**(1), 44–55.

Fung, W. and Hsieh, D.A., Will hedge funds regress towards index-like products? *J. Invest. Manage.*, 2007, **5**(32), 46–65.

Giamouridis, D. and Paterlini, S., Regular(ized) hedge fund clones. *J. Financ. Res.*, 2010, **33**(3), 223–247.

Gilli, M. and Kellezi, E., The threshold accepting heuristic for index tracking. In *Financial Engineering, E-Commerce, and Supply Chain*, pp. 1–18, 2009 (Kluwer: Dordrecht).

Gilli, M. and Winker, P., Heuristic optimization methods in econometrics. In *Handbook of Computational Econometrics*, edited by D. Beasley and E. Kontoghiorghes, pp. 81–120, 2009 (Wiley).

Giuzio, M., Genetic algorithm versus classical methods in sparse index tracking. *Decis. Econ. Finance*, 2017, **40**(1-2), 243–256.

Giuzio, M and Paterlini, S., Un-diversifying during crises: Is it a good idea?. *Comput. Manage. Sci.*, 2018. https://doi.org/10.1007/s10287-018-0340-y

Giuzio, M., Eichhorn-Schott, K., Paterlini, S. and Weber, V., Tracking hedge funds using sparse clones. *Ann. Oper. Res.*, 2018, **266**(1–2), 349–371.

Golosnoy, V., Gribisch, B. and Seifert, M.I., Exponential smoothing of realized portfolio weights. *J. Empir. Finance*, 2019, **53**, 222–237.

Gu, S., Kelly, B. and Xiu, D., Empirical asset pricing via machine learning. *Rev. Financ. Stud.*, 2020, **33**, 2223–2273. https://doi.org/10.1093/rfs/hhaa009

Hastie, T., Tibshirani, R. and Friedman, J., *The Elements of Statistical Learning - Data Mining,Inference and Prediction*, 2nd ed. 2001 (Springer: Stanford, CA).

Kremer, P.J., Talmaciu, A. and Paterlini, S., Risk minimization in multi-factor portfolios: What is the best strategy?. *Ann. Oper. Res.*, 2018, **266**(1–2), 255–291.

Kremer, P.J., Lee, S., Bogdan, M. and Paterlini, S., Sparse portfolio selection via the sorted ℓ_1-norm. *J. Bank. Finance*, 2020, **110**(6B), 1–41.

Krink, T., Mittnik, S. and Paterlini, S., Differential evolution and combinatorial search for constrained index tracking. *Ann. Oper. Res.*, 2009, **172**(1), 153–176.

Mainik, G., Mitov, G. and Rüschendorf, L., Portfolio optimization for heavy-tailed assets: Extreme risk index vs. markowitz. *J. Empir. Finance*, 2015, **32**(2), 115–134.

Malkiel, B.G., Returns from investing in equity mutual funds 1971 to 1991. *J. Finance.*, 1995, **50**(2), 549–572.

Rudolf, M., Wolter, H.-J. and Zimmermann, H., A linear model for tracking error minimization. *J. Bank. Finance*, 1999, **23**(1), 85–103.

Santosa, F. and Symes, W.W., Linear inversion of band-limited reflection seismograms. *SIAM J. Sci. Stat. Comput.*, 1986, **7**(4), 1307–1330.

Sorenson, E., Miller, K. and Samak, V., Allocating between active and passive management. *Financ. Anal. J.*, 1998, **54**(5), 18–31.

Su, W. and Candès, E.J., SLOPE is adaptive to unknown sparsity and asymptotically minimax. *Ann. Stat.*, 2016, **44**(3), 1038–1068.

Tibshirani, R., Regression shrinkage and selection via the LASSO. *R. Stat. Soc.*, 1996, **58**(1), 267–288.

Weston, J., Elisseeff, A. and Schoelkopf, B., Use of the zero-norm with linear models and kernel methods. *J. Mach. Learn. Res.*, 2003, **3**, 1439–1461.

Xing, X., Hub, J. and Yang, Y., Robust minimum variance portfolio with ℓ_∞ constraints. *J. Bank. Finance*, 2014, **46**, 107–117.

Yen, Y.-M., Sparse weighted norm minimum variance portfolio. *Rev. Financ.*, 2015, **20**(3), 1259–1287.

Yen, Y.-M. and Yen, T.-J., Solving norm constrained portfolio optimization via coordinate-wise descent algorithms. *Comput. Stat. Data Anal.*, 2014, **76**(1), 737–759.

Appendices

Appendix 1. Alternative penalty functions

A.1. *Smoothly clipped absolute deviation*

To resolve the problem of large biased coefficient values of the LASSO, Fan and Li (2001) developed the non-convex smoothly clipped absolute deviation (SCAD) penalty, given by:

$$\rho_\lambda(\boldsymbol{w}) = \sum_{i=1}^{K} \left\{ \lambda |w_i| \times \mathbb{1}(|w_i| \leq \lambda) + \frac{-w_i^2 + 2a\lambda|w_i| - \lambda^2}{2(a-1)} \times \mathbb{1}(\lambda < |w_i| \leq a\lambda) + \frac{(a+1)\lambda^2}{2} \times \mathbb{1}(a\lambda < |w_i|) \right\} \tag{A1}$$

where a is a threshold parameter, and $\mathbb{1}(\cdot)$ is the indicator function, which is equal to one when the argument in the parenthesis is true and zero otherwise.

The SCAD consists of three regions that, depending on the estimated coefficients, determine the value of $\rho_\lambda(\boldsymbol{w})$. As long as $|w_i| < \lambda$, the SCAD increases linearly and thus has the same shrinkage abilities as the LASSO. Furthermore, as we start to impose a larger lambda parameter, this linear region will expand and the penalty gets closer to the LASSO. On the other hand, given a large estimated coefficient, that is $|w_i| > a\lambda$, the SCAD imposes an upper bound on the value of the penalty function. Even if the estimated coefficient value increases past this point, it will no longer inflate the penalty function. Consequently, the SCAD has the tendency to produce extreme positive and negative weights, when compared to the LASSO. Still, the maximum attainable value of such coefficients is limited by the added budget constraint, as a larger weight for asset i only goes in hand with a lower weight for asset j. In between the two extreme points, and when $\lambda < |w_i| \leq a\lambda$, the penalty 'smoothly clips' the estimated parameters.

From an economic perspective and given the correlation structure, the SCAD penalizes those assets which receive a small weight and thus probably have less explanatory power. On the other hand, weights which exceed the threshold of $a\lambda$ are considered to be important predictors and are not intended to be further penalized (Fastrich *et al.* 2015).†

A.2. *Logarithmic penalty*

Besides the SCAD, we also consider the non-convex logarithmic penalty function (LOG) (see e.g. Weston *et al.* 2003), given by:

$$\rho_\lambda(\boldsymbol{w}) = \lambda \times \sum_{i=1}^{K} (\log(|w_i| + \gamma) - \log(\gamma)) \tag{A2}$$

where $0 < \gamma < 1$ is a constant to avoid the occurrence of an undefined logarithm, when $w_i = 0$. The LOG penalty can be considered as an approximation to the ℓ_0 or cardinality constraint. Furthermore, it closely approximates the behavior of the ℓ_q-Norm with $0<q<1$,‡ which even intensifies when $q \to 0$. Compared to the ℓ_q -Norm, however the LOG penalty leads to sparser solutions and has shown to possess good sparsity recovery properties (see i.e. Candes *et al.* 2008, Giuzio *et al.* 2018).

From an economic perspective, an increase of a small weight is penalized more heavily, than an associated increase of a large weights. Based on the correlation structure, it thus promotes selecting only a few large coefficients with high explanatory power, while disregarding small and unnecessary coefficients.

Appendix 2. Proof of theorems

A.3. *Proof of theorem 2.1*

Proof Clearly, permuting columns of $\boldsymbol{R}$ and switching their signs correspond, respectively, to permuting and switching signs of the coefficients of the solution. Let $\widetilde{\boldsymbol{R}}$ denote the matrix $\boldsymbol{R}$ after multiplying its columns by the signs of corresponding coefficients of $\hat{\boldsymbol{w}}$ and after permuting them with respect to the order of magnitudes of $\hat{\boldsymbol{w}}$'s coefficients. It holds $rank(\widetilde{\boldsymbol{R}}) = k$ and it is enough to prove that the solution, $\widetilde{\boldsymbol{w}}$, to the problem with modified design matrix satisfies the claim. We have $\widetilde{\boldsymbol{w}}_1 \geq \cdots \geq \widetilde{\boldsymbol{w}}_K \geq 0$ and hence $\widetilde{\boldsymbol{w}}$ solves

$$\arg\min_{\boldsymbol{w}} \frac{1}{2} \|\boldsymbol{Y} - \tilde{\boldsymbol{R}}\boldsymbol{w}\|_2^2 + \lambda^\top \boldsymbol{w} \quad \text{s.t.} \quad \boldsymbol{w}_1 \geq \cdots \geq \boldsymbol{w}_K \geq 0 \tag{A3}$$

The claim is true for $k = K$. Suppose that $k < K$ and denote by N the null space of $\widetilde{\boldsymbol{R}}$. Then we have $dim(N) = K - k > 0$. We need to show that at least $K - k$ inequalities defining the feasible set in (A3) become equalities for $\widetilde{\boldsymbol{w}}$. Suppose that this is not true and that we have a partition of $\{1, \ldots, K\}$ into two sets of indices, I_{eq} and I_{ineq}, such that

$$i \in I_{eq} \implies \tilde{\boldsymbol{w}}_i = \tilde{\boldsymbol{w}}_{i+1}, \quad i \in I_{ineq} \implies \tilde{\boldsymbol{w}}_i > \tilde{\boldsymbol{w}}_{i+1},$$

† For our empirical investigations, we choose $a = 3$.

‡ For applications of this norm in the area of index tacking, see i.e. Fernholtz *et al.* (1998), Fastrich *et al.* (2014), Chen *et al.* (2013), Giuzio (2017).

$$|I_{eq}| < K - k, \tag{A4}$$

with a convention that $\widetilde{\boldsymbol{w}}_{K+1} := 0$. Now, consider the vector subspace H, defined as $H := \{w \in \mathbb{R}^K | w_i = w_{i+1},\ i \in I_{eq}\}$, once again with the convention that $w_{K+1} := 0$. Since we have less than $K - k$ linear equations defining the subspace, it holds $dim(H) > k$. Therefore, there exists the non-zero vector, d, such as $d \in N \cap H$ and without the loss of generality we can assume that $\lambda^\mathsf{T} d \le 0$. We can also find $\delta > 0$ such that for all $i \in I_{ineq}$ it holds $\widetilde{\boldsymbol{w}}_i + \delta \cdot d_i > \widetilde{\boldsymbol{w}}_{i+1} + \delta \cdot d_{i+1}$. Consider the vector defined as $c := \widetilde{\boldsymbol{w}} + \delta \cdot d$. Since $c \in H$, we have $c_i = c_{i+1}$, for $i \in I_{eq}$, and the construction of δ yields $c_i > c_{i+1}$, for $i \in I_{ineq}$. Consequently, c is a feasible point in the optimization problem (A3). Now,

$$\begin{aligned}\frac{1}{2}\left\|\boldsymbol{Y} - \tilde{\boldsymbol{R}}c\right\|_2^2 + \lambda^\mathsf{T} c &= \frac{1}{2}\left\|\boldsymbol{Y} - \tilde{\boldsymbol{R}}\tilde{\boldsymbol{w}} - \delta\tilde{\boldsymbol{R}}d\right\|_2^2 + \lambda^\mathsf{T}\tilde{\boldsymbol{w}} + \delta\cdot\lambda^\mathsf{T} d \\ &\le \frac{1}{2}\left\|\boldsymbol{Y} - \tilde{\boldsymbol{R}}\tilde{\boldsymbol{w}}\right\|_2^2 + \lambda^\mathsf{T}\tilde{\boldsymbol{w}},\end{aligned} \tag{A5}$$

By the construction we have $c \ne \widetilde{\boldsymbol{w}}$, hence the last inequality contradicts either the optimality of $\widetilde{\boldsymbol{w}}$ (if '$<$' holds) or the uniqueness of the solution (if '$=$' holds). ■

A.4. *Proof of theorem 2.2*

Proof Suppose that $\hat{\boldsymbol{w}}_i > \hat{\boldsymbol{w}}_{i+1}$ and define $\psi := e_{i+1} - e_i$, where e_j is an indicator vector of index j. There exists $\varepsilon > 0$ such that the vector $c_\delta := \hat{\boldsymbol{w}} + \delta \cdot \psi$ is feasible for any $\delta \in (0, \varepsilon)$. Now, $\hat{\boldsymbol{w}}$ is the solution to the minimization problem with the objective $f(w) := \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{w}\|_2^2$ and under the constraints $\boldsymbol{w}_1 \ge \cdots \ge \boldsymbol{w}_K \ge 0$. Since f is differentiable, the directional derivative of f exists for any argument and any direction. We have

$$\nabla_\psi f(\hat{\boldsymbol{w}}) := \lim_{\delta\to 0}\frac{f(\hat{\boldsymbol{w}} + \delta\cdot\psi) - f(\hat{\boldsymbol{w}})}{\delta} = \lim_{\substack{\delta\to 0^+ \\ \delta\in(0,\varepsilon)}}\frac{f(c_\delta) - f(\hat{\boldsymbol{w}})}{\delta} \ge 0,$$

where the inequality follows from the optimality of $\hat{\boldsymbol{w}}$ and the construction of c_δ. Now, deriving the gradient of f in $\hat{\boldsymbol{w}}$ yields

$$\nabla f(\hat{\boldsymbol{w}}) = -\boldsymbol{R}^\mathsf{T}(\boldsymbol{Y} - \boldsymbol{R}\hat{\boldsymbol{w}}) + \lambda$$

and, since it always holds $\nabla_\psi f(\hat{\boldsymbol{w}}) = \psi^\mathsf{T}\nabla f(\hat{\boldsymbol{w}})$, we get $(e_i - e_{i+1})^\mathsf{T}\boldsymbol{R}^\mathsf{T}(\boldsymbol{Y} - \boldsymbol{R}\hat{\boldsymbol{w}}) \ge (e_i - e_{i+1})^\mathsf{T}\lambda$. Now, the condition $\hat{\boldsymbol{w}}_i > \hat{\boldsymbol{w}}_{i+1}$ implies

$$\begin{aligned}\lambda_i - \lambda_{i+1} &\le \left(\boldsymbol{R}_i - \boldsymbol{R}_{i+1}\right)^\mathsf{T}\left(\boldsymbol{Y} - \boldsymbol{R}\hat{\boldsymbol{w}}\right) \\ &= \left(\boldsymbol{R}_i - \boldsymbol{R}_{i+1}\right)^\mathsf{T}\left(r_P - \boldsymbol{R}_i\hat{\boldsymbol{w}}_i - \boldsymbol{R}_{i+1}\hat{\boldsymbol{w}}_{i+1}\right) \\ &= \left(\boldsymbol{R}_i - \boldsymbol{R}_{i+1}\right)^\mathsf{T} r_P \\ &\quad - \left(d^2\hat{\boldsymbol{w}}_i + \boldsymbol{R}_i^\mathsf{T}\boldsymbol{R}_{i+1}\hat{\boldsymbol{w}}_{i+1} - \boldsymbol{R}_{i+1}^\mathsf{T}\boldsymbol{R}_i\hat{\boldsymbol{w}}_i - d^2\hat{\boldsymbol{w}}_{i+1}\right) \\ &= \left(\boldsymbol{R}_i - \boldsymbol{R}_{i+1}\right)^\mathsf{T} r_P - \left(d^2 - \boldsymbol{R}_i^\mathsf{T}\boldsymbol{R}_{i+1}\right)\left(\hat{\boldsymbol{w}}_i - \hat{\boldsymbol{w}}_{i+1}\right) \\ &\le \left(\boldsymbol{R}_i - \boldsymbol{R}_{i+1}\right)^\mathsf{T} r_P,\end{aligned} \tag{A6}$$

after applying the Cauchy-Schwarz inequality, which ends the proof. ■