Sparse spanning portfolios and under-diversification with second-order stochastic dominance

Stelios Arvanitis, Olivier Scaillet, Nikolas Topaloglou

February 2023

Abstract

We develop and implement methods for determining whether relaxing sparsity constraints on portfolios improves the investment opportunity set for risk-averse investors. We formulate a new estimation procedure for sparse second-order stochastic spanning based on a greedy algorithm and Linear Programming. We show the optimal recovery of the sparse solution asymptotically whether spanning holds or not. From large equity datasets, we estimate the expected utility loss due to possible under-diversification, and find that there is no benefit from expanding a sparse opportunity set beyond 30 assets. The optimal sparse portfolio invests in 10 industry sectors with a larger weighting on small size, high book-to-market, and momentum stocks from the S&P 500 index and cuts tail risk when compared to a sparse mean-variance portfolio. On a rolling-window basis, the number of assets shrinks to 10 assets in crisis periods.

Keywords and phrases: Nonparametric estimation, stochastic dominance, spanning, under-diversification, greedy algorithm, Linear Programming.

JEL Classification: C13, C14, C44, C58, C61, D81, G11.

[a] Athens University of Economics and Business, [b] University of Geneva and Swiss Finance Institute: corresponding author; GFRI, Bd du Pont d'Arve 40, 1211 Geneva, Switzerland; +41 22 379 88 16; olivier.scaillet@unige.ch, [c] IPAG Business School and Athens University of Economics and Business.

1 Introduction

We know for decades that the diversification benefits measured by the volatility of portfolio returns are limited when we invest beyond 10 to 20 assets; see e.g. Evans and Archer (1968), Klemkosky and Martin (1975), Elton and Gruber (1977). Practitioners coin the term over-diversification. At the opposite end of the spectrum, we often observe under-diversification among households (Campbell (2006), Calvet, Campbell and Sodini (2007)). It might be caused by information acquisition costs (Van Nieuwerburgh and Veldkamp (2010)), overconfidence (Anderson (2013)), solvency requirements (Liu (2014)), or overweighting low probability events (Dimmock et al. (2021)). A characteristic-based demand system might also explain why institutions and households hold a small set of stocks (Koijen and Yogo (2019)).

Possible over-diversification contributes to motivating the recent literature on sparse construction of mean-variance (MV) portfolios within the Modern Portfolio Theory (Markowitz (1952)) through imposing constraints on the portfolio weights; see e.g. Jagannathan and Ma (2003), DeMiguel et al. (2009), Brodie et al. (2009), Fan, Zhang and Yu (2012), Ao, Li and Zheng (2019), and Caner, Medeiros and Vasconscelos (2021). Such a construction limits the impact of transaction costs, and eases monitoring and risk management. It also achieves statistical regularisation of the investment portfolio in the presence of ill-conditioned large covariance matrices. Whether limitations of diversification benefits beyond a given small number of assets still hold true when we leave the MV paradigm is an open problem. This paper targets the following questions: Is it possible to build a sparse portfolio of dimension q from a large set of assets of dimension p so that we cannot get further improvement from considering additional assets in a second-order stochastic dominance (SSD) paradigm? If not, how much do we loose by limiting ourselves to this sparse portfolio in terms of expected utilities compatible with SSD? Can we design an optimization algorithm to compute this sparse portfolio from available data? Do we have the asymptotic statistical guarantee that we cannot improve on the estimated expected utility loss due to under-diversification by considering another sparse portfolio of the same fixed dimension?

The theory of stochastic dominance (SD) gives a systematic framework for analyzing investor behavior under uncertainty (see Chapter 4 of Danthine and Donaldson (2014) for an introduction oriented towards finance). Stochastic dominance ranks portfolios based on general regularity conditions for decision making under risk (Hadar and Russell (1969), Hanoch and Levy (1969), and Rothschild and Stiglitz (1970)). SD uses a distribution-free assumption framework which allows for nonparametric statistical estimation and inference methods. We can see SD as a flexible model-free alternative to MV dominance of Modern Portfolio Theory (Markowitz (1952)). The MV criterion is consistent with Expected Utility for elliptical distributions such as the normal distribution (Chamberlain (1983), Owen and Rabinovitch (1983), Berk (1997)) but has limited economic meaning when we cannot completely characterize the probability distribution by its location and scale. Simaan (1993), Athayde and Flores (2004), and Mencia and Sentana (2009) develop a mean-variance-skewness framework based on generalizations of elliptical distributions that are fully characterized by their first three moments. SD presents a further generalization that accounts for all moments of the return distributions without necessarily assuming a particular family of distributions.

Second-order SD (SSD) spanning (Arvanitis et al. (2019)) is a model-free alternative to MV spanning of Huberman and Kandel (1987) (see also Jobson and Korkie (1989), De Roon, Nijman, and Werker (2001)). Spanning occurs if introducing new securities or relaxing investment constraints does not improve the investment possibility set for a given class of investors. MV spanning checks if the MV frontier of a set of assets is identical to the MV frontier of a larger set made of those assets plus additional assets (Kan and Zhou (2012), Penaranda and Sentana (2012)). Here we investigate such a problem for investors with risk-averse preferences which are interested in the whole return distributions generated by two sets of assets, a sparse subset of dimension q (10 assets) with a limited number of assets coming from a much larger set of dimension p (500 stocks). We take the viewpoint of a fund manager making available a fund invested in a potentially restricted number of assets to all risk averse investors. We do not take the viewpoint of a wealth manager making available a dedicated portfolio taylor made to the specific utility of her customer in a privare bank or family office.

The first contribution of the paper is to introduce the concept of sparse SSD spanning. We propose a theoretical measure for sparse spanning based on second-order stochastic dominance. For economic interpretation, we provide with a representation based on a class of concave utility functions without assuming differentiability. When sparse SSD spanning occurs, a risk-averse investor will not improve her expected utility by shifting from the sparse subset to the larger investment opportunity set. On the contrary, if it does not occur, the risk-averse investor will suffer an expected utility loss since we work with a subset instead of the full set of assets. Hence we further provide a lower bound that takes the interpretation of an optimal utility loss that cannot be improved upon by any sparse subset made of q assets. We know that we suffer a loss because of the sparsity constraint but we cannot do better through investing optimally in only q assets under an SSD criterion. To check sparse SSD spanning on data, we develop consistent and feasible estimation procedures based on Linear Programming (LP) and a greedy algorithm, namely the Forward Stepwise algorithm. We use a finite set of increasing piecewise-linear functions, restricted to the bounded empirical supports, that are constructed as convex mixtures of appropriate "ramp functions" (in the spirit of Russel and Seo (1989)) in our representation as in Arvanitis, Scaillet and Topaloglou (2020a,b). For every such utility function, we solve two embedded linear maximization problems. It is an improvement over the implementation in Arvanitis and Topaloglou (2017) and Arvanitis, Scaillet and Topaloglou (2020b) where they formulate the empirical counterpart in terms of Mixed-Integer Programming (MIP) problems. MIP problems are NP-complete, and far more difficult to solve. Our numerical approximations are simple and fast since they are based on standard LP. They suit better computationally intensive optimisation methods, which otherwise become quickly computationally demanding in empirical work on large data sets, especially when relying on resampling techniques to compute confidence intervals. Those formulations are reminiscent of the LP programs developed in the early papers of testing for

SSD efficiency of a given portfolio by Post (2003) and Kuosmanen (2004) (see also Scaillet and Topaloglou (2010)).

Since we aim at a sparse solution computed from a large dimensional problem, we rely on a greedy optimisation algorithm. We use a discrete combinatorial algorithm for maximizing a function subject to a cardinality constraint. It starts with the empty set, and then adds elements to it in r iterations. In each iteration, the algorithm adds to its current solution the single element increasing the value of this solution by the most, i.e., the element with the largest marginal value with respect to the current solution. In the context of submodular maximization (see Buchbinder and Feldman (2018) for a survey), this simple Forward Stepwise algorithm checking for incremental gain at each step using nested models is usually referred to simply as "the greedy algorithm". In the case of submodular functions, it selects a set of a given cardinality (10 assets) to produce a numerical solution that is provably within a constant factor of the optimum given by the optimal set of the same cardinality (Nemhauser, Wolsey and Fisher (1978)), and it turns out to be the best approximation ratio possible for the problem (Nemhauser and Wolsey (1978)). A submodular function has a natural diminishing return property: adding an element to a larger set results in smaller marginal increase in the value of the function compared to adding the element to a smaller set. Bian et al. (2017) extend guarantee results of the greedy algorithm for cardinality constrained maximization of non-submodular nondecreasing set functions, in particular nondecreasing standard LP problems with non-degenerate basic feasible solution (Bertsimas and Tsitsiklis (1997), Ch. 3) that we implement in our empirics.

We choose that approach over penalization methods currently used for building sparse MV portfolios for two reasons. First, we wish to bound the relative error without any assumptions on the underlying sparsity for the true parameter. It is useful to show the consistency of our empirical strategy irrespective of sparse spanning being present or absent. Our proof relies on the recent work of Elenberg et al. (2018) (see Das and Kempe (2011) for the linear regression case). Contrary to prior work in the MV setting, we require neither assumptions on the sparsity of the underlying problem nor i.i.d. returns. Rather, we exploit their result establishing multiplicative approximation guarantees from the best-case sparse solution. Our results improve over previous work by providing bounds on a solution that is guaranteed to match the desired sparsity and cannot be further decreased. Convex methods for linear regressions such as the standard LASSO objective (Tibshirani (1996)) require strong assumptions on the model and the data, such as the unrepresentable condition on the parameter vector and i.i.d. data (Zhao and Yiu (2006), Meinshausen and Buhlmann (2006)), in order to provide exact sparsity guarantees on the recovered solution (see Zhang (2009) for use of these assumptions in greedy least squares regression). More specifically, when the number r of iterations is equal to $r = q \ln T$, T being the time-series sample size, we show that the algorithm provides a consistent estimate of the bound of the expected utility loss computed from financial returns satisfying a mixing condition. Mixing holds true for many time series models such as ARMA models as well as several GARCH and stochastic volatility processes (see Francq and Zakoian (2011) for several examples). It allows us to build a path of the estimated bound as a function of the sparsity constraint q, and verify when we have a sufficiently large q to get sparse SSD spanning, namely when the bound vanishes. Second, the only input we need is the sparsity number q of assets. Hence, we avoid the selection problem of a tuning parameter, namely the regularization parameter in penalization methods. As discussed in Brodie et al. (2009), a portfolio selection with a LASSO approach regulates the amount of shorting. In our setting, we use short-sales constraints which corresponds to using an implicit large regularisation parameter for the LASSO penalty. Our numerical approach based on a greedy algorithm however does not require the true portfolio to be sparse, and a large regularisation parameter is not required for developing valid statistical inference. As a by-product, our approach also provides a selection algorithm for sparse MV spanning under multivariate normality using the equivalence with sparse SSD spanning for elliptical distributions. It allows to bypass the regularization of ill-conditioned estimates of large covariance matrices (see e.g. Fan, Liao, and Shi (2015), Ledoit and Wolf (2017)).

The second contribution of the paper aims at checking on large data sets of equity returns whether sparse SSD holds or not.

(...)The paper is organized as follows. In Section 2, we establish our probabilistic framework, and review the definition of SSD. In Section 3, we define the relevant concept of sparse SDD spanning and provide with convenient functional representations. We discuss the concept of approximate sparse spanning in Section 4. Given a fixed support dimension, it specifies the low dimensional portfolio set that comes closer (in an appropriate sense defined later on in the paper) to span the high dimensional one. In Section 5, we construct an estimate of the bound for sparse SSD spanning by using empirical analogues. We exploit the limiting distribution of the empirical process underlying the estimator which has the form of a Gaussian process. Our estimation strategy builds on LP and a Forward Stepwise algorithm. We show the asymptotic optimal recovery of the sparse solution, namely statistical approximation guarantee of the greedy algorithm output for a given q when T becomes large. In Section 6, we describe the numerical implementation aspects of our empirical procedures. We perform a Monte Carlo experiment in Section 7. In Section 8, we analyze large data sets of equity returns to study whether sparse SSD holds or not and compare with results given by the construction of sparse MV portfolios with the MAXSER approach of Ao, Li and Zheng (2019). We provide concluding remarks in Section 9. We provide our proofs in the Appendix.

2 Background-Second Order Stochastic Dominance

We describe our limiting economy for a large number of financial assets. We denote the financial returns by a process X^{∞} living in $\ell^{\infty}(\mathbb{N}, \mathbb{R})$, which is the space of bounded real valued sequences equipped with the uniform metric. X_i denotes the i^{th}, $i \in \mathbb{N}$ coordinate, X denotes the projection of X^{∞} in the first p coordinates, and $\mathbb{P}$ denotes the distribution of X^{∞}. We suppress dependence on p for brevity. F denotes the cdf of the distribution of the random vector X.

We introduce the associated portfolio weights with short-sales constraints. Short-sales constraints on the asset allocation promote sparsity (Brodie et al. (2009)); our approach can be used to trace further patterns of (desired) sparsity. The set Λ_∞ is a non-empty subset of the $\mathbb{N}$-simplex $\left\{\lambda \in \mathbb{R}^{\mathbb{N}} : \lambda_i \geq 0, i \in \mathbb{N}, \sum_{i=0}^{\infty} \lambda_i = 1\right\}$, and for $p \in \mathbb{N}$, $\Lambda = \left\{\lambda \in \Lambda_\infty, \sum_{i=0}^{p-1} \lambda_i = 1\right\}$ denotes the $p-1$ dimensional unit sub-simplex of Λ_∞ and K is a non-empty closed subset of Λ. In the present context, X is a random vector of financial returns for p base assets, while Λ represents a set of portfolios formed on X. The process X^∞ idealizes the high dimensional situation in the limiting case where $p \to \infty$. Our first assumption specifies probabilistic properties for X^∞. It requires mild moment existence conditions (bounded sequence of first order moments), and a lower bound on the associated supports consistent with non-logarithmic returns. Here, $\bar{co}$ denotes the closure of the convex hull. Given the restrictions that its elements satisfy, Λ_∞ is considered topologized by the l_1 norm, and λ, κ denote generic elements of Λ^∞.

Assumption 1. $\max_{0<i\leq+\infty} \mathbb{E}\left[|X_i|\right] < +\infty$. $Z := \bar{co}\left[\cup_i supp\left(X_i\right)\right]$ *and* $\inf Z > -\infty$.

In this context, for any $\lambda \in \Lambda_\infty$, $\sum_{i=0}^{\infty} \lambda_i X_i$ is a well defined random variable since due to the monotonicity of the integral, $\mathbb{E}\left[|\sum_{i=0}^{\infty} \lambda_i X_i|\right] \leq \max_i \mathbb{E}\left[|X_i|\right] < +\infty$. It implies that $\mathbb{E}\left[\left(z - \sum_{i=0}^{\infty} \lambda_i X_i\right)_+\right]$ is continuous in z, λ via dominated convergence, and that it is also bounded in λ for any z, even though Λ_∞ is not (l_1-) totally bounded. Along with the Lipschitz continuity property of $(\cdot)_+$, it also implies that for any $\lambda \neq \kappa$,

$$\sup_{z\in\mathbb{R}} \left| \mathbb{E}\left[\left(z - \sum_{i=0}^{\infty} \kappa_i X_i\right)_+\right] - \mathbb{E}\left[\left(z - \sum_{i=0}^{\infty} \lambda_i X_i\right)_+\right] \right| \leq \max_i \mathbb{E}\left[|X_i|\right] \sum_{i=0}^{\infty} \left(\kappa_i + \lambda_i\right), \quad (1)$$

i.e. the Lower Partial Moment Differential (LPMD) $D\left(z, \kappa, \lambda, \mathbb{P}\right) := \mathbb{E}\left[\left(z - \sum_{i=0}^{\infty} \kappa_i X_i\right)_+\right] - \mathbb{E}\left[\left(z - \sum_{i=0}^{\infty} \lambda_i X_i\right)_+\right]$ is also bounded and continuous in z, λ, κ. Assumption 1 thus facilitates the definition of Second order Stochastic Dominance (SSD) for the constructed portfolios:

Definition 1. κ SSD dominates λ, written $\kappa \underset{\text{SSD}}{\succeq} \lambda$, iff $D(z, \kappa, \lambda, \mathbb{P}) \leq 0$ for all $z \in Z$.

The definition is simply an adaptation of the usual SSD relation in our high dimensional framework. Using the classical Russell and Seo (1989) utility representations, we obtain the well-known result that $\kappa \underset{\text{SSD}}{\succeq} \lambda$ iff the former is preferred to the latter by every increasing and concave utility. Thus, SSD exemplifies universal choices w.r.t. every insatiable and risk averse investor.

3 Sparse SSD Spanning

Arvanitis et al. (2019) define the notion of SSD Spanning as an extension of the Mean-Variance analogue. It involves comparison of portfolio sets that are not necessarily singletons.

Definition 2. $K \underset{\text{SSD}}{\succeq} \Lambda$ iff $\forall \lambda \in \Lambda$, $\exists \kappa \in K : \kappa \underset{\text{SSD}}{\succeq} \lambda$.

If the sets are not related by inclusion and $K \underset{\text{SSD}}{\succeq} \Lambda$ then we have necessarily $K \underset{\text{SSD}}{\succeq} K \cup \Lambda$. Furthermore, spanning would be trivial if $K \supseteq \Lambda$ were allowed. Hence, we can always consider that K lies inside Λ. Spanning admits an interesting economic interpretation precisely when $K \subseteq \Lambda$; it means that extension of the investment possibility set from K to Λ does not improve investment opportunities for any risk averter. Hence, no spanning means that the extension contains a non dominated element. This is formalized as follows: $K \underset{\text{SSD}}{\not\succeq} \Lambda$ iff $\exists \lambda \in \Lambda : \forall \kappa \in K$, $\kappa \underset{\text{SSD}}{\not\succeq} \lambda$, i.e. λ is maximal (efficient) w.r.t. K.

Under some further structure on K, namely compactess, SSD spanning admits an empirically useful characterization involving a saddle-type point of the LPMDs.

Lemma 1. *Under Assumption 1, and if moreover K is compact, then $K \underset{SSD}{\succeq} \Lambda$ iff $\sup_{\Lambda} \inf_{K} \sup_{z \in Z} D(z, \kappa, \lambda, \mathbb{P}) \leq 0$.*

We extend the notion in the high dimensional setting, by also allowing a potentially unknown low dimensional investment opportunity set to SSD span a high dimensional superset. In order to formally define this and extend it to the limiting case where $p \to \infty$, we introduce the following notation for the support of a portfolio set: $\mathrm{csupp}\,(K) := \#\{i : \kappa_i \neq 0, \kappa \in K\}$. By construction $\mathrm{csupp}\,(\Lambda) = p$. We suppose that as $p \to +\infty$ and $\lim_{p\to\infty} \Lambda = \Lambda_\infty$, where the limit is interpreted in the Painleve-Kuratowski convergence mode. The sequence $(\Lambda)_p$ is by construction monotone increasing.

Definition 3 (Sparse Spanning SSD)**.** For a fixed q, there exists a $K \subset \Lambda$ with $\mathrm{csupp}\,(K) \leq q$ and such that $K \underset{\text{SSD}}{\succeq} \Lambda$.

Definition 3 generalizes Definition 2 in a twofold manner; First it allows for a limiting high dimensional setting thus providing the proper framework for addressing the empirical finance questions that were presented in the Introduction. Second, it only prescribes the existence of a "low-dimensional" spanning subset of Λ, whereas for the original definition the spanning subset is exogenously given. It implies that any procedure designed to test whether SS-SSD holds, would have to search for a spanning set inside the collection of "low-dimensional" subsets of Λ. It is useful even in the case where SS-SSD does not hold. As the following paragraph suggests, such a procedure, if consistent (what is shown below), would end up with a sparse portfolio set that "comes as close as possible" to SSD span its high dimensional universe of portfolios.

As in Lemma 1, we obtain a useful characterization of SS-SSD by assuming some further topological structure on the portfolio weights sets. Consider the collection $\mathcal{L}_{p,q} = \{K \subset \Lambda : K \text{ closed}, 0 < \mathrm{csupp}\,(K) \leq q\}$. When Λ is itself a simplicial complex, then $\mathcal{L}_{p,q}$ is also a simplicial complex of dimension $q-1$. Then and if $p \geq 2q$, $\mathcal{L}_{p,q}$ has a geometric realization as a sub-simplex of the standard $p-1$ simplex (see the Geometric Realization Theorem in Edelsbrunner (2014)).

Lemma 2. *Under Assumption 1, suppose moreover that Λ is closed in the Euclidean topology. Then, SS-SSD is equivalent to that for large enough p, and fixed $q < p$, $\inf_{\mathcal{L}_{p,q}} \sup_\Lambda \inf_K \sup_{z\in Z} D(z, \kappa, \lambda, \mathbb{P}) \leq 0$. The latter is equivalent to that $\inf_{\mathcal{L}_{\infty,q}} \sup_{\Lambda_\infty} \inf_{K^\star} \sup_{z\in Z} D(z, \kappa, \lambda, \mathbb{P}) \leq 0$.*

The possibility of interchanging the order of appearance of the optimization operators in the characterization $\inf_{\mathcal{L}_{p,q}} \sup_{\Lambda} \inf_K \sup_{z \in Z} D(z, \kappa, \lambda, \mathbb{P}) \leq 0$ to $\sup_{z \in Z} \sup_{\Lambda} \inf_{\mathcal{L}_{p,q}} \inf_K$ will greatly facilitate numerical aspects as well as the derivations of limiting properties for the empirical procedures. It actually holds via the use of appropriate minimax theorems and the extension of our assumption framework.

Lemma 3. *Suppose that Assumptions 1 and 3 hold, and that Λ is closed in the Euclidean topology. Then,* for all p,

$$\inf_{\mathcal{L}_{p,q}} \sup_{\Lambda} \inf_{K} \sup_{z \in Z} D(z, \kappa, \lambda, \mathbb{P}) = \sup_{z \in Z} \sup_{\Lambda} \inf_{\mathcal{L}_{p,q}} \inf_{K} D(z, \kappa, \lambda, \mathbb{P}).$$

Lemma 3 eases the numerical optimization implementation. For an arbitrary threshold z, $\sup_{\Lambda} \inf_{\mathcal{L}_{p,q}} \inf_K D(z, \kappa, \lambda, \mathbb{P}) = \inf_{\mathcal{L}_{p,q}} \inf_K \mathbb{E}\left[\left(z - \sum_{i=0}^{\infty} \kappa_i X_{t,i}\right)_+\right] - \inf_{\Lambda} \mathbb{E}\left[\left(z - \sum_{i=0}^{\infty} \lambda_i X_{t,i}\right)_+\right]$, so that we can separate the optimizations w.r.t. the "parameter sets" Λ and $\mathcal{L}_{p,q} \times K$. It is useful especially in the case where we approximate the outer optimization over Z by some discretization, as is usually the case in empirical numerical implementations.

3.1 Approximate Sparse Spanning

We consider the optimization problem $M(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}) := \sup_{z \in Z} \sup_{\Lambda} \inf_{\mathcal{L}_{p,q}} \inf_K D(z, \kappa, \lambda, \mathbb{P})$. Even if $M(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}) > 0$, so that there exists no K with $\operatorname{csupp}(K) \leq q$ for which SS-SSD holds, any solution to this problem has an interpretation as an approximate sparse spanning subset of Λ in the sense of an expected utility loss as stated in the next proposition. For $\mathcal{P}(Z)$ the set of probability distributions (or equivalently cdfs) supported on Z, and for any F there, consider the Russell-Seo increasing and concave utility (see Russell and Seo (1989)) $u_F(x) := \int_Z \min(0, x - z)\, dF(z)$.

Proposition 1. $K \in \mathcal{L}_{p,q}$ *does not solve* $\sup_{\Lambda} \sup_{z \in Z} \inf_{\mathcal{L}_{p,q}} \inf_K D(z, \kappa, \lambda, \mathbb{P})$*, iff there exists some $\lambda \in \Lambda$ and some u_F such that* $\mathbb{E}\left[u_F\left(\sum_{i=0}^{\infty} \lambda_i X_i\right)\right] - \mathbb{E}\left[u_F\left(\sum_{i=0}^{\infty} \kappa_i X_i\right)\right] > M(\Lambda, \mathcal{K}_{p,q}, \mathbb{P})$ *for any $\kappa \in K$.*

Hence, $M(\Lambda, \mathcal{L}_{p,q}, \mathbb{P})$ is the optimal expected utility difference that the elements of any sparse subset of Λ of support dimension equal to q can achieve w.r.t. the elements of Λ uniformly over the Russell-Seo utilities. The solutions to $\sup_{\Lambda} \sup_{z \in Z} \inf_{\mathcal{L}_{p,q}} \inf_{K} D(z, \kappa, \lambda, \mathbb{P})$ are those subsets that actually achieve this optimality bound. Lemma 3 along with the monotonicity of (Λ_p) implies also that as $p \to \infty$, $M(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}) \to M(\Lambda_{\infty}, \mathcal{L}_{\infty,q}, \mathbb{P})$.

3.2 Sparse Approximately Efficient Elements

For any $z \in Z$, due to the Russell-Seo utilities set up (see Russell and Seo (1989)) and the utility representations in Proposition 1, any solution to $\inf_{\Lambda} \mathbb{E}\left[\left(z - \sum_{i=0}^{\infty} \lambda_i X_{t,i}\right)_+\right]$ is an efficient element of Λ since it must be non-dominated as an optimizer of the utility that corresponds to F that concentrates its mass on z. In this respect, any portfolio that results from the solution to $\inf_{\mathcal{L}_{p,q}} \inf_{K} \mathbb{E}\left[\left(z - \sum_{i=0}^{\infty} \kappa_i X_{t,i}\right)_+\right]$ must be a sparse element of Λ of support at most q. That sparse element optimally approximates the efficient element since $\sup_{\Lambda} \mathbb{E}\left[u_F\left(\sum_{i=0}^{\infty} \lambda_i X_i\right)\right] - \sup_{\mathcal{L}_{p,q}} \sup_{K} \mathbb{E}\left[u_F\left(\sum_{i=0}^{\infty} \kappa_i X_i\right)\right]$ is less than or equal to $\sup_{\Lambda}\left[u_F\left(\sum_{i=0}^{\infty} \lambda_i X_i\right)\right] - \mathbb{E}\left[u_F\left(\sum_{i=0}^{\infty} \kappa_i X_i\right)\right]$, for any κ of support at most q. It is also an efficient element of maximizer over $\mathcal{L}_{p,q}$ of $\sup_{K} \mathbb{E}\left[u_F\left(\sum_{i=0}^{\infty} \kappa_i X_i\right)\right]$. When $M(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}) \leq 0$, the solution to $\inf_{\mathcal{L}_{p,q}} \inf_{K} \mathbb{E}\left[\left(z - \sum_{i=0}^{\infty} \kappa_i X_{t,i}\right)_+\right]$ is also efficient in Λ. When $K \in \mathcal{L}_{p,q}$ maximizes $\sup_{K} \mathbb{E}\left[u_F\left(\sum_{i=0}^{\infty} \kappa_i X_i\right)\right]$ uniformly in z, but does not span Λ, then there necessarily exist efficient elements of Λ that are not in K. Then, the portfolio that solves $\sup_{K} \mathbb{E}\left[u_F\left(\sum_{i=0}^{\infty} \kappa_i X_i\right)\right]$ uniformly in z, is by construction an efficient element of K that minimizes $\mathbb{E}\left[u_F\left(\sum_{i=0}^{\infty} \lambda_i X_i\right)\right] - \mathbb{E}\left[u_F\left(\sum_{i=0}^{\infty} \kappa_i X_i\right)\right]$ uniformly w.r.t. the efficient set of Λ and the Russell-Seo utilities. Interestingly, it is an efficient element of K that maximizes a utility that corresponds to a distribution F that concentrates its mass on some threshold z.

As $p \to \infty$, any accumulation point of the solution to $\inf_{\mathcal{L}_{p,q}} \inf_{K} \mathbb{E}\left[\left(z - \sum_{i=0}^{\infty} \kappa_i X_{t,i}\right)_+\right]$ is a q-sparse approximate efficient element of Λ_{∞}. If it is unique and independent of z, then it is also a portfolio bound for the set of q-sparse portfolios (for the concept of portfolio bounds on finite dimensional portfolio spaces; see Arvanitis, Post and Topaloglou (2021)). In this case, every efficient element of Λ_∞ is approximated by the same q-sparse approximate efficient element of Λ_∞. If $\inf_{\Lambda_\infty} \mathbb{E}\left[(z - \sum_{i=0}^{\infty} \lambda_i X_{t,i})_+\right]$ has also a unique solution independent of z, then this is also a portfolio bound (of potentially infinite support) of Λ_∞. When SS-SSD holds and q is large enough, then those two bounds coincide, and thereby Λ_∞ admits a q-sparse bound.

4 Sparse Optimization: Greedy Algorithm and Statistical Guarantees

In this section, and given the latency of $D(z, \kappa, \lambda, \mathbb{P})$, we are interested in the empirical approximation of the element of $\mathcal{L}_{p,q}$ that approximately spans Λ for a fixed q. We employ the empirical analogues of the functionals above that characterize spanning, and design the sparse optimization via a greedy algorithm. We establish consistency using the results on statistical guarantees by Elenberg et al. (2018). We derive the usual parametric $\sqrt{T}$ rate and the limiting distribution, and construct a conservative inferential procedure based on (fast) subsampling.

Consider the sequence $(X_t^\infty)_{t\in\mathbb{Z}}$, where for all t, $X_t^\infty \underset{\mathrm{d}}{=} X^\infty$ and $\underset{\mathrm{d}}{=}$ denotes equality in distribution. Suppose, that for some p, a sample of $(X_t)_{t=1,\dots,T}$ is available from the sequence $(X_t)_{t\in\mathbb{Z}}$. Denote with $\mathbb{P}_T$ its empirical distribution function in $\mathbb{R}^p$ (in what follows $\mathbb{P}$ also identifies the distribution of X_0 in $\mathbb{R}^p$ without inconsistency due to the Daniel-Kolmogorov Theorem). We approximate $D(z, \kappa, \lambda, \mathbb{P})$ by $D(z, \kappa, \lambda, \mathbb{P}_T)$ and design a procedure that evaluates $\inf_{\mathcal{L}_{p,q}} \sup_\Lambda \inf_K \sup_{z\in Z} D(z, \kappa, \lambda, \mathbb{P}_T)$.

Given Lemma 3, we design our empirical procedure as follows: for fixed q, formulate the empirical optimization problem $\sup_{z\in Z}\sup_{\Lambda}\inf_{\mathcal{L}_{p,q}}\inf_{K} D(z,\kappa,\lambda,\mathbb{P}_T)$, as

$$\begin{aligned} M(\Lambda,\mathcal{L}_{p,q},\mathbb{P}_T) &:= \sup_{z\in Z}\left[\mathcal{K}(\Lambda,\mathcal{L}_{p,q},z,\mathbb{P}_T)-\mathcal{L}(\Lambda,z,\mathbb{P}_T)\right], \\ \mathcal{K}(\Lambda,\mathcal{L}_{p,q},z,\mathbb{P}_T) &:= \inf_{\mathcal{L}_{p,q}}\inf_{K}\tfrac{1}{T}\textstyle\sum_{t=0}^{T}\left(z-\sum_{i=0}^{\infty}\kappa_i X_{t,i}\right)_+ , \\ \mathcal{L}(\Lambda,z,\mathbb{P}_T) &:= \inf_{\Lambda}\tfrac{1}{T}\textstyle\sum_{t=0}^{T}\left(z-\sum_{i=0}^{\infty}\lambda_i X_{t,i}\right)_+ . \end{aligned} \tag{2}$$

Given $\mathcal{K}(\Lambda,\mathcal{L}_{p,q},z,\mathbb{P}_T)$ the numerical technology that evaluates $\mathcal{L}(\Lambda,z,\mathbb{P}_T)$ and $M(\Lambda,\mathcal{L}_{p,q},\mathbb{P}_T)$ is the same with the one employed in the SD literature in the low dimensional settings. The main issue here is to design a procedure that evaluates $\mathcal{K}(\Lambda,\mathcal{L}_{p,q},z,\mathbb{P}_T)$. The outer optimization there involves searching over low dimensional subsets of Λ. As explained in the introduction, we favor a procedure based on a greedy algorithm that approximates $\mathcal{K}(\Lambda,\mathcal{L}_{p,q},z,\mathbb{P}_T)$ over procedures based on penalization. We use the Forward Stepwise algorithm (Algorithm 2 in Elenberg et al. (2018)). Let us denote $r_T(q)$ the number of iterations performed.

Algorithm 1. *Forward Stepwise Selection (see p. 3542 of Elenberg et al. (2018)). Inputs: the sparsity Parameter $q < p$, the # of iterations $r_T(q)$, for a given set S the set function $2^p \to \mathbb{R}$ defined as*

$$func(S) := \inf_{csupp(S)\leq q}\frac{1}{T}\sum_{t=0}^{T}\left(z-\sum_{i=0}^{\infty}\kappa_i X_{i,t}\right)_+ .$$

a Choose the initial set S_0,

b for $i = 1,\ldots,r_T(q)$ do,

c $s := \arg\max_{j\in[p]/S_{i-1}} func(S_{i-1}\cup\{j\}) - func(S_{i-1})$,

d $S_i := S_{i-1}\cup\{s\}$.

The last step (d), for $i = r_T(q)$, returns $\mathcal{K}^{\mathrm{FS}}(\Lambda,\mathcal{L}_{p,q},z,\mathbb{P}_T,r_T(q))$ namely the numerical approximation of $\mathcal{K}(\Lambda,\mathcal{L}_{p,q},z,\mathbb{P}_T)$ in (2) by the greedy algorithm. The next section provides with details on the numerical aspects of the three optimizations appearing in (2), including the implementation of FSS.

We now examine the issues of consistency, rates of convergence and limiting distribution of $M(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}_T)$ given $\mathcal{K}^{\text{FS}}(\Lambda, \mathcal{L}_{p,q}, z, \mathbb{P}_T, r_T(q))$. In order to study these, we use an assumption concerning a property of restricted strong convexity (see Ch. 9 of Wainwright (2019) for reviewing restricted strong convexity in high-dimensional statistics) and smoothness for the LPMs as $p \to \infty$. In this section, we also denote Λ with Λ_p whenever it is important to keep track of the dimension of the portfolio space. For $p \gg m \in \mathbb{N}$, we denote the set $\{(\lambda, \lambda^\star) \in \Lambda_p \times \Lambda_p : \text{csupp}(\lambda) \leq m, \text{csupp}(\lambda^\star) \leq m, \text{csupp}(\lambda - \lambda^\star) \leq m\}$ with $\Lambda_{(m)}$. $\tilde{\Lambda}_{(m)}$ denotes the set obtained by keeping the first component λ of the pairs $(\lambda, \lambda^\star)$ that define the elements of $\Lambda_{(m)}$.

Assumption 2. *[Restricted Strong Convexity-Restricted Smoothness (RSC/RS)] X has a continuous density f. $\mathbb{E}\left(z - \sum_{i=0}^{\infty} \kappa_i X_{0,i}\right)_+$ is twice differentiable w.r.t. any κ appearing in some pair of $\Lambda_{(\lfloor q(\ln(T+1)) \rfloor)}$ for all $z \in Z$. For $m_{\lfloor q(\ln(T+1)) \rfloor}$ denoting the supremum and $M_{\lfloor q(\ln(T+1)) \rfloor}$ the infimum over $\Lambda_{(\lfloor q(\ln(T+1)) \rfloor)}$, of the smallest and the largest eigenvalues of the Hessian matrix of $\mathbb{E}\left(z - \sum_{i=0}^{\infty} \kappa_i X_{0,i}\right)_+$, we have that as $T \to \infty$, $\frac{m_{\lfloor q(\ln(T+1)) \rfloor}}{M_{\lfloor q(\ln(T+1)) \rfloor}} \ln T \to +\infty$ uniformly in Z.*

Let us characterize that assumption on an example, which shows that this assumption is mild. For the $\lfloor q(\ln(T+1)) \rfloor$-dimensional, due to Assumption 1, Theorem 1 of Savare (1996) and given the distributional derivative of $(x)_+$ (see p.1 in Savare (1996)), we obtain that $\mathbb{E}\left(z - \sum_{i=0}^{\infty} \kappa_i X_{0,i}\right)_+$ is twice differentiable and the Hessian assumes the form $\int_{\mathbb{R}^q} XX^T \delta\left(z - \kappa^T X\right) f(X)\, dX$, where δ denotes the Dirac Delta function. Using Example 27 in Estrada and Kanwal (2012), the latter equals $C_\kappa \int_{z - \kappa^T X} XX^T f(X)\, dX$, for a constant $C_\kappa > 0$ that depends on κ and emerges in the surface measure of the hyperplane $z - \kappa^T X$. Suppose now that f is a normal density. Then, using the results of Cong et al. (2017) (see their Algorithm 2), we have that the Hessian takes the form $C_{z,\kappa} \mathbb{E}\left(\left[\left(\text{Id}_q - \frac{1}{\Delta} V \kappa \kappa^T\right) X + \frac{1}{\Delta} V \kappa z\right] \left[\left(\text{Id}_q - \frac{1}{\Delta} V \kappa \kappa^T\right) X + \frac{1}{\Delta} V \kappa z\right]^T\right)$, where V is the second moment matrix of X, $C_{z,\kappa} > 0$ is an integration constant depending on both z and κ and $\Delta = \kappa^T V \kappa$. A simple calculation, along with the constraint $z = \kappa^T X$, yields that the Hessian equals $C_{z,\kappa}V$. Given that $\frac{\sup_{z,\kappa} m}{\inf_{z,\kappa} M} \geq \sup_{z,\kappa} \frac{m}{M}$, Assumption 2 follows if $\frac{\text{Condition Number of } V}{\ln T} \to 0$, i.e., if the condition number of V is dominated by $\ln T$ as $T \to \infty$. It means that we can also accommodate a slowly diverging condition number. The analysis in Par. 5 of Kim and Pollard (1990) implies that the same condition suffices when f is continuously differentiable, which is more inline with the bounded support framework of our applications. When V has a Kac-Murdock-Szego type Toeplitzian structure (Trench (1999)), where $V_{i,j} = v^{|i-j|}$, $i, j = 1, \ldots, p$ for $v \in [0, 1)$, Assumption 2 holds trivially since the condition number is then uniformly bounded in p (Trench (1999), p. 182). Such matrices appear in zero mean normalised autoregressive progresses. In the case of the zero mean spiked identity model (Example 7.18 in Wainwright (2019)), where $V = \mathbf{Id} + \mu\mathbf{1}\mathbf{1}'$ for some $\mu \in [0, 1)$, the results in the aforementioned example imply that Assumption 2 holds when μ converges to zero with T. The condition number asymptotic restriction above does not hold when μ is strictly positive and fixed, a situation however that can accommodate RSC conditions in temporally i.i.d. Gaussian frameworks (Example Theorem 7.16 of Wainwright (2019)). When X is not necessarily zero mean, then the variational representations of the maximum eigenvalue $\lambda_{\max}(A)$, and the minimum eigenvalue $\lambda_{\max}(A)$, of a pd matrix A, imply that Assumption 2 holds whenever $\frac{\text{Condition Number of Var}(X) + \text{Condition Number of } \mathbb{E}(X)\mathbb{E}(X)'}{\ln T} \to 0$ or $\frac{\text{Condition Number of Var}(X)}{\ln T} + \frac{\lambda_{\max}(\mathbb{E}(X)\mathbb{E}(X)')}{\lambda_{\min}(\text{Var}(X)) \ln T} \to 0$.

The RSC/RS assumption along with Assumption 3 imply analogous RSC/RS properties for the empirical LPMs $\frac{1}{T}\sum_{t=0}^{T} \left(z - \sum_{i=0}^{\infty} \kappa_i X_{t,i}\right)_+$ with probability converging to one (w.h.p.). It enables the use of the results of Elenberg et al. (2018) on statistical guarantees for the Forward Stepwise algorithm.

Our analysis also depends on the asymptotic behavior of the empirical processes $\sqrt{T}D\left(z, \kappa, \lambda, \mathbb{P}_T - \mathbb{P}\right)$, $G_T\left(z, \kappa, \lambda\right) := \sqrt{T}\left[g\left(z, \lambda, \mathbb{P}_T\right) - g\left(z, \lambda, \mathbb{P}\right)\right]^T (\kappa - \lambda)$, and of the empirical moment process $\frac{1}{T}\sum_{t=0}^{T} \left(z_T - \sum_{i=0}^{\infty} \lambda_i X_{t,i}\right)_+$, where $g\left(z, \lambda, Q\right) := \mathbb{E}_Q\left[X\mathbb{I}\left\{z \geq \sum_{i=0}^{\infty} \boldsymbol{\lambda}_i X_i\right\}\right]$ and $\mathbb{E}_Q$ denotes integration w.r.t. the measure Q. Specifically, consistency is facilitated if the first and the second processes are asymptotically tight over appropriate subsets of parameters, and the third process (locally) uniformly converges to its population counterpart. This behavior depends on stationarity and mixing rates for the returns process involved as well as a stricter moment existence condition compared to Assumption 1.

Assumption 3. *$(X_t^{\infty})_{t\in\mathbb{Z}}$ is strictly stationary and absolutely regular with mixing coefficients $(\beta_m)_{m\in\mathbb{N}}$ that satisfy $\beta_m \sim b^m$ for some $b \in (0,1)$, as $m \to \infty$, and $\max_{0<i\leq+\infty} \mathbb{E}\left[|X_i|^{2+\varepsilon}\right] < +\infty$ for some $\varepsilon > 0$.*

The stationarity, ergodicity and mixing rates conditions as well as the moment existence condition hold for several geometrically ergodic (finite dimensional), linear as well as GARCH type models with values in Euclidean spaces. Those are frequently employed in empirical finance with data consistent parameter restrictions-see Francq and Zakoian (2011). Using the Daniell-Kolmogorov Theorem we have that stationarity and mixing rates hold for the $(X_t^{\infty})_{t\in\mathbb{Z}}$ process whenever they hold uniformly over the collection of finite dimensional parts of the process. Thereby, they hold whenever the finite dimensional parts of the process are consistent with the aforementioned models with appropriately uniform parameter restrictions.

We obtain the following limit theory; let $\ell^{\infty}(Z \times \Lambda^{\infty} \times \Lambda^{\infty})$ denote the space of real valued bounded functions on $Z \times \Lambda^{\infty} \times \Lambda^{\infty}$ equipped with the sup norm. We use $\rightsquigarrow$ to denote weak convergence.

Theorem 1. *Suppose that Assumptions 1, 3 hold. Then, (a) $\frac{1}{T}\sum_{t=0}^{T}\left(z_T - \sum_{i=0}^{\infty}\lambda_i X_{t,i}\right)_+ \rightsquigarrow \mathbb{E}\left[\left(z - \sum_{i=0}^{\infty}\lambda_i X_{t,i}\right)_+\right]$, for any $z, z_T \in Z$ with $z_T \to z$, and uniformly in Λ^{∞}. Furthermore, suppose also that $\frac{\ln p}{\sqrt{T}} \to 0$. Then as $T \to \infty$, with $\kappa \in \tilde{\Lambda}_{(\lfloor q(\ln T+1)\rfloor)}$ (b) $\sqrt{T}D\left(z,\kappa,\lambda,\mathbb{P}_T - \mathbb{P}\right) \rightsquigarrow \mathcal{G}(z,\kappa,\lambda)$, in $\ell^{\infty}(Z \times \Lambda^{\infty} \times \Lambda^{\infty})$, where $\mathcal{G}(z,\kappa,\lambda)$ is a zero mean Gaussian process with covariance kernel defined by*

$$\mathcal{V}\left[(z_1,\kappa_1,\lambda_1),(z_2,\kappa_2,\lambda_2)\right] := \sum_{t\in\mathbb{Z}} Cov\left[\mathcal{I}(z_1,\kappa_1,\lambda_1,X_0),\mathcal{I}(z_2,\kappa_2,\lambda_2,X_t)\right] \quad , \tag{3}$$

where $\mathcal{I}(z,\kappa,\lambda,X_t) := \left(z-\sum_{i=0}^{\infty}\kappa_i X_{t,i}\right)_+ - \left(z-\sum_{i=0}^{\infty}\lambda_i X_{t,i}\right)_+$. *Finally, (c)*

$$\limsup_{T\to\infty} \mathbb{E}[\sup_{z} \sup_{\Lambda_{(\lfloor q(\ln T+1)\rfloor)}} G_T(z,\kappa,\lambda)] < \infty.$$

The condition $\frac{\ln p}{\sqrt{T}} \to 0$ that appears in the final pair of results of the theorem is somewhat stricter than the usual $\frac{\ln p}{T} \to 0$ that appears in the literature, it however facilitates standard rates and limiting Gaussianity for the empirical processes involved and thus the results that go beyond consistency. It ensures that the bracketing entropy of Λ grows at an appropriate rate in order for tightness to hold in the limit. For the notion of the bracketing entropy numbers of a metric space, see Section 5 of Andrews (1994) and Ch. 2 of van der Vaart and Wellner (1996). In our context, it corresponds to the mapping that keeps track of the logarithm of the minimal number of δ-brackets (w.r.t. the l_1 norm) of real sequences with absolutely convergent series needed to cover the particular neighborhood, for each $\delta > 0$.

Using the above, we obtain the following consistency result.

Theorem 2. *Suppose that Assumptions 1, 2, 3, hold, that* Λ *is closed and for large enough* p *it is also convex, and that* $\frac{\ln p}{\sqrt{T}} \to 0$. *For fixed* q, *as* $T\to\infty$, *and uniformly in* Z,

$$\mathcal{K}^{FS}(\Lambda,\mathcal{L}_{p,q},z,\mathbb{P}_T,q\ln T) \rightsquigarrow \mathcal{K}(\Lambda^{\infty},\mathcal{L}_{\infty,q},z,\mathbb{P}). \tag{4}$$

Consequently, $M^{FS}(\Lambda,\mathcal{L}_{p,q},\mathbb{P}_T,q\ln T) \rightsquigarrow M(\Lambda_{\infty},\mathcal{L}_{\infty,q},\mathbb{P})$, *where* $M^{FS}(\Lambda,\mathcal{L}_{p,q},\mathbb{P}_T,r_T(q)) := \sup_{z\in Z}\left[\mathcal{K}^{FS}(\Lambda,\mathcal{L}_{p,q},z,\mathbb{P}_T,r_T(q)) - \mathcal{L}(\Lambda,z,\mathbb{P}_T)\right]$.

Whenever Assumption 2 holds for some $q^{\star}\in\mathbb{N}$, Theorem 2 implies then that the mapping $q \to M^{\mathrm{FS}}(\Lambda,\mathcal{K}_{p,q},\mathbb{P}_T,q\ln T)$ converges in probability to $q \to M(\Lambda_{\infty},\mathcal{K}_{\infty,q},\mathbb{P})$ uniformly in $q \le q^{\star}$.

Theorem 2 holds whether we have sparse spanning or not at the limit. We do not need to assume sparsity in the population. The statistical guarantee result of Theorem 2 is a strong advantage of the greedy algorithm over penalization methods.

We are further occupied with the determination of the rates of convergence and the distributional limit for the deviation $M^{\mathrm{FS}}\left(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}_T, q\left(\ln T\right)^2\right) - M\left(\Lambda^{\infty}, \mathcal{L}_{\infty,q}, \mathbb{P}\right)$, that gauges the gap between $M^{\mathrm{FS}}\left(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}_T, q\left(\ln T\right)^2\right)$, which is returned by the greedy algorithm on the data, and the limit $M\left(\Lambda^{\infty}, \mathcal{L}_{\infty,q}, \mathbb{P}\right)$. To this end, we augment r_T to $q\left(\ln T\right)^2$, in order to facilitate arguments that approximate the infimum of $\frac{1}{T}\sum_{t=0}^{T}\left(z - \sum_{i=0}^{\infty}\kappa_i X_{t,i}\right)_+$ over the empirical solution in $\mathcal{L}_{p,q}$, by the the infimum of $\mathbb{E}\left[\left(z - \sum_{i=0}^{\infty}\kappa_i X_{t,i}\right)_+\right]$ over the population solution. Given the second result of Theorem 1, we obtain standard rates and a distributional limit defined as a saddle type point of a zero mean Gaussian process, using among others the generalized Delta method (see Fang and Santos (2019)).

Theorem 3. *Suppose that Assumptions 1, 3, 2 hold, that Λ is closed and for large enough p it is also convex, and that $\frac{\ln p}{\sqrt{T}} \to 0$. Suppose furthermore that the following conditions hold: i) (Condition CO) the mapping $z \to \mathbb{E}\left[\left(z - \sum_{i=0}^{\infty}\kappa_i X_{t,i}\right)_+\right]$ is strictly concave for any κ with $csupp\left(\kappa\right) \leq q$, and ii) (Condition CM) for any $z > \inf Z$, $\mathbb{E}\left[\left(z - \sum_{i=0}^{\infty}\lambda_i X_{t,i}\right)_+\right]$ has a compact subset of minimizers over Λ_{∞} and $\mathbb{E}\left[\left(z - \sum_{i=0}^{\infty}\kappa_i X_{t,i}\right)_+\right]$ has a compact set of minimizers of support less than or equal to q. Then as $T \to \infty$,*

$$\sqrt{T}\left(M^{FS}\left(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}_T, q\left(\ln T\right)^2\right) - M\left(\Lambda_{\infty}, \mathcal{L}_{\infty,q}, \mathbb{P}\right)\right) \rightsquigarrow \sup \inf_{(z,\lambda,\kappa)\in\Gamma} \mathcal{G}\left(z, \lambda, \kappa\right), \quad (5)$$

where $\mathcal{G}\left(z, \lambda, \kappa\right)$ is a zero mean Gaussian process with covariance kernel defined by

$$\mathcal{V}\left[\left(z_1, \lambda_1, \kappa_1\right), \left(z_2, \lambda_2, \kappa_2\right)\right] := \sum_{t\in\mathbb{Z}} Cov\left[\mathcal{I}\left(z_1, \lambda_1, \kappa_1, X_0\right), \mathcal{I}\left(z_2, \lambda_2, \kappa_2, X_t\right)\right] \quad , \quad (6)$$

$\mathcal{I}\left(z, \lambda, \kappa, X_t\right)$ as in Theorem 1, and $\Gamma := \arg\max_{z\in Z, \lambda\in\Lambda_{\infty}} \min_{csupp(\kappa)\leq q} D\left(z, \lambda, \kappa, \mathbb{P}\right)$.

Since Λ_{∞} is separable, there are no measurability problems with the definition of $\sup\inf_{(z,\lambda,\kappa)\in\Gamma} \mathcal{G}\left(z, \lambda, \kappa\right)$. Furthermore, for any $\lambda^{\star} \in \Lambda_{\infty}$, $\mathcal{G}_{\lambda^{\star}}\left(z, \kappa\right) = \mathcal{G}\left(z, \lambda^{\star}, \kappa\right)$. Regarding CO, $\mathbb{E}\left(z - \sum_{i=0}^{\infty}\kappa_i X_{0,i}\right)_+$ is twice differentiable w.r.t. z and the second derivative assumes the form $\int_{\mathbb{R}^q} \delta\left(z - \kappa^T X\right) f\left(X\right) dX$. It is equal to the probability attributed by f to the hyperplane $z = \sum_{i=0}^{\infty} \kappa_i X_i$ which is positive if X has a non degenerate covariance matrix. Under normality, the condition is thus guaranteed by the aforementioned limiting behavior on V that also guarantees Assumption 2. For CM, Theorem 4.5 of Beer and Lucchetti (1991) says that compactness of the set of minimizers is a generic property in the sense of Baire category. Hence, it is expected to hold at least for a dense subset of Z, due to Assumption 1 and dominated convergence. The exclusion of the trivial threshold from the considerations is innocuous since $\mathcal{G}(\inf Z, \lambda, \kappa)$ is identically zero. CO and CM implies that $\Gamma - \{\inf Z\} \times \Lambda_\infty \times \tilde{\Lambda}_{(q)}$ is compact and thereby the generalized Delta method is applicable.

Theorem 3 allows for the construction of a feasible inferential procedure based on subsampling in the spirt of Linton et al. (2014) (see also Linton et al. (2005)) that approximates the asymptotic quantiles of the limit in (5). To get a viable numerical strategy, we design the subsampling technique to avoid the costly numerical search of the Forward Stepwise algorithm inside each subsample. To this end, let $\kappa_{z,T}$ denote the solution of $\inf_{\text{csupp}(\kappa) \leq q} \frac{1}{T} \sum_{t=0}^{T} \left(z_t - \sum_{i=0}^{\infty} \kappa_i X_{t,i}\right)_+$ over $\mathcal{L}_{p,q}$. Denote with $\Gamma^\star$ the subset of Γ that contains the triplets at which some accumulation point of $\kappa_{z,T}$ appears. Let $0 < b_T \leq T$, and consider the subsamples from the original observations $(X_j)_{j=t,\ldots t+b_T-1}$ for all $t = 1, 2, \ldots, T - b_T + 1$. For $\alpha \in (0, 1)$, denote with $q_{T,B_T}(1-\alpha)$ the $1-\alpha$ quantile of the subsample empirical distribution of $\left(\sqrt{b_T}\left(\sup_{Z \times \Lambda_p} D\left(z, \kappa_{z,T}, \lambda, \mathbb{P}_{t,b_T}\right) - M^{\text{FS}}\left(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}_T, q\left(\ln T\right)^2\right)\right)\right)_{t=1,\ldots,T-b_T+1}$, where $\mathbb{P}_{t,b_T}$ denotes the empirical distribution of $(X_j)_{j=t,\ldots t+b_T-1}$ and we use the same $\kappa_{z,T}$ across subsamples. Hence, we get a fast subsampling method (Hong and Scaillet (2006)).

Our final result depends on a condition on the elements of $\Gamma^\star$ that avoids limiting degeneracies (Condition ND below). They would imply poor higher-order properties for the conservative inference that we consider in Proposition 2. We say that a triple in $\Gamma^\star$ is trivial if the variance of $\mathcal{G}$ there is zero. We have triviality when the first element of the triple is $\inf Z$. It is also the case when λ coincides with the κ appearing in the triple. Then, λ is by construction an efficient element of Λ_∞ that is also q-sparse. Whenever the elements of X_p are linearly independent for p larger than the maximum desired value of q for the analysis at hand, trivialities can occur only if SS-SSD holds. This linear independence holds for the Gaussian case that exemplifies Assumption 2 above.

Proposition 2. *Suppose that (Condition ND) for the given* q, $\Gamma^\star$ *contains at least one non trivial triplet. Under the premises of Theorem 3, if* $b_T \to \infty$, $\frac{b_T}{T} \to 0$ *and* $\alpha < \frac{1}{2}$, *then we get the conservative result:*

$$\limsup_{T\to\infty} \mathbb{P}\left[M\left(\Lambda_\infty, \mathcal{L}_{\infty,q}, \mathbb{P}\right) \in \left(M^{FS}\left(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}_T, q\left(\ln T\right)^2\right) \mp q_{T,B_T}\left(1-\alpha\right)\right)\right] \geq 1-\alpha. \quad (7)$$

If moreover there exists a unique q*-sparse element of* Λ *that appears in every triple in* $\Gamma^\star$, *then we get the exact result:*

$$\lim_{T\to\infty} \mathbb{P}\left[M\left(\Lambda_\infty, \mathcal{L}_{\infty,q}, \mathbb{P}\right) \in \left[M^{FS}\left(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}_T, q\left(\ln T\right)^2\right) \mp q_{T,B_T}\left(1-\alpha\right)\right]\right] = 1-\alpha. \quad (8)$$

Under linear independence, ND would hold whenever every q-sparse efficient element is matched by an efficient element of appropriately large support compared to the maximum desired level of q for the underlying analysis.

The evaluation of the quantile by subsampling has small computational burden since we avoid the costly sparse optimization w.r.t. κ inside each subsample. We only need to compute once $\kappa_{z,T}$ on the full sample and keep it fixed across subsamples. Usually, Z is approximated by some finite discretization and optimization w.r.t. λ is performed via linearization of the SD conditions and the use of LP methods. LP formulations coupled with fast subsampling avoids the computational cost of sparse optimization, and the asymptotic results in (7)-(8) hold as long as the discretized set converges to a dense subset of Z.

In the special case where $M\left(\Lambda_\infty, \mathcal{L}_{\infty,q}, \mathbb{P}\right)$ has a unique solution, (5) implies asymptotic normality. A unique solution is possible whenever SS-SSD does not hold. It occurs whenever the maximal expected utility difference between an efficient element of Λ_∞ and its approximate counterpart of dimension q occurs at a unique Russell-Seo utility for a unique pair of efficient-approximate efficient portfolios. In such a case, we can exploit normality to obtain a result like (8). A feasible normality result requires a consistent estimator for the limiting variance. It can be obtained via a subsampling methodology that does not involve subsample optimizations, as long as stricter moment conditions hold for X_0, and a non-degeneracy condition for the covariance kernel of $\mathcal{G}$ holds in some neighborhood of the optimizer.

5 Numerical Implementation

Let us now describe how we can implement our spare SSD spanning approach in practice with LP and a greedy algorithm. For $q < p$, we consider the following empirical optimization problem

$$\sup_{z\in Z}\sup_{\Lambda}\inf_{\mathcal{L}_{p,q}}\inf_{K} D\left(z,\kappa,\lambda,\mathbb{P}_T\right). \tag{9}$$

The utility class interpretation of Arvanitis, Scaillet and Topaloglou (2020a,b) implies that we can represent (9) in terms of expected utility as:

$$\sup_{\boldsymbol{\lambda}\in\Lambda;u\in\mathcal{U}}\inf_{\mathcal{L}_{p,q}}\inf_{\kappa\in\mathrm{K}} \mathbb{E}_{\mathbb{P}}\left[u\left(X^{\mathrm{T}}\boldsymbol{\lambda}\right)-u\left(X^{\mathrm{T}}\boldsymbol{\kappa}\right)\right], \tag{10}$$

with $\mathcal{U} := \left\{u\in\mathcal{C}^0 : u(y)=\int_{\underline{x}}^{\overline{x}}v(x)r(y;x)dx\ v\in\mathcal{V}\right\}$, $\mathcal{V} := \left\{v:\mathcal{X}\to\mathbb{R}_+ : \int_{\mathcal{X}}v\left(x\right)=1\right\}$, and $r(y;x) := (y-x)1(y\leq x),\ (x,y)\in\mathcal{X}^2$.

The set $\mathcal{U}$ in (10) is comprised of normalized, increasing, and concave utility functions that are constructed as convex mixtures of elementary Russell et Seo (1989) ramp functions $r(y;x),\ x\in\mathcal{X}$. This representation is used in the numerical implementation via

$$\sup_{u\in\mathcal{U}}\inf_{\mathcal{L}_{p,q}}\left(\sup_{\boldsymbol{\lambda}\in\Lambda}\mathbb{E}_{\mathbb{P}_T}\left[u\left(X^{\mathrm{T}}\boldsymbol{\lambda}\right)\right]-\sup_{\boldsymbol{\kappa}\in\mathrm{K}}\mathbb{E}_{\mathbb{P}_T}\left[u\left(X^{\mathrm{T}}\boldsymbol{\kappa}\right)\right]\right). \tag{11}$$

In (11), we approximate every element of $\mathcal{U}$ with arbitrary prescribed accuracy using a finite set of increasing and concave piecewise-linear functions in the following way.

Let N_1, N_2 denote integers greater than or equal to 2. First, $\mathcal{X}$ is partitioned into N_1

equally spaced values as $\underline{x} = z_1 < \cdots < z_{N_1} = \overline{x}$, where $z_n := \underline{x} + \frac{n-1}{N_1-1}(\overline{x} - \underline{x})$, $n = 1, \cdots, N_1$. Second, $[0, 1]$ is partitioned as $0 < \frac{1}{N_2-1} < \cdots < \frac{N_2-2}{N_2-1} < 1$. Using these partitions, consider:

$$\sup_{u \in \underline{\mathcal{U}}} \left(\sup_{\boldsymbol{\lambda} \in \Lambda} \mathbb{E}_{\mathbb{P}_T} \left[u \left(X^{\mathrm{T}} \boldsymbol{\lambda} \right) \right] - \sup_{\boldsymbol{\kappa} \in \mathrm{K}} \mathbb{E}_{\mathbb{P}_T} \left[u \left(X^{\mathrm{T}} \boldsymbol{\kappa} \right) \right] \right), \tag{12}$$

with $\quad \underline{\mathcal{U}} := \left\{ u \in \mathcal{C}^0 : u(y) = \sum_{n=1}^{N_1} v_n r(y; z_n) \; v \in V \right\}$,

and $\quad V := \left\{ v \in \left\{ 0, \frac{1}{N_2-1}, \cdots, \frac{N_2-2}{N_2-1}, 1 \right\}^{N_1} : \sum_{n=1}^{N_1} v_n = 1 \right\}$.

By construction, every $u \in \underline{\mathcal{U}}$ consists of at most N_2 linear line segments with end-points at N_1 possible outcome levels. Furthermore, $\underline{\mathcal{U}} \subset \mathcal{U}$, which is finite as it has $N_3 := \frac{1}{(N_1-1)!} \prod_{i=1}^{N_1-1} (N_2 + i - 1)$ elements and $\underline{\eta_T}$ approximates η_T from below as the partitioning scheme is refined ($N_1, N_2 \to \infty$). Then, for every $u \in \underline{\mathcal{U}}$, the two embedded maximization problems in (12) can be solved using LP. Consider $c_{0,n} := \sum_{m=n}^{N_1} (c_{1,m+1} - c_{1,m}) z_m$, $c_{1,n} := \sum_{m=n}^{N_1} w_m$, and $\mathcal{N} := \{n = 1, \cdots, N_1 : v_n > 0\} \bigcup \{N_1\}$. Then, for any given $u \in \underline{\mathcal{U}}$, $\sup_{\boldsymbol{\lambda} \in \Lambda} \mathbb{E}_{\mathbb{P}_T} \left[u \left(X^{\mathrm{T}} \boldsymbol{\lambda} \right) \right]$ is the optimal value of the objective function of the following LP problem in canonical form: $\max T^{-1} \sum_{t=1}^{T} y_t$ s.t. $y_t - c_{1,n} X_t^{\mathrm{T}} \boldsymbol{\lambda} \leq c_{0,n}$, $t = 1, \cdots, T$, $n \in \mathcal{N}$, $\sum_{i=1}^{M} \lambda_i = 1$, $\lambda_i \geq 0$, $i = 1, \cdots, M$, and y_t free, $t = 1, \cdots, T$. The LP problem always has a feasible solution and has $\mathcal{O}(T + N)$ variables and constraints. In the empirical application, we take $N_1 = 10$ and $N_2 = 5$. Thus, we end up with $N_3 = \frac{1}{9!} \prod_{i=1}^{9} (4 + i) = 715$ distinct utility functions and $2N_3 = 1\,430$ small LP problems, which is perfectly manageable with modern-day computer hardware and solver software. We use a desktop PC with a 3.6 GHz, 24-core Intel i7 processor, with 128 GB of RAM, using MATLAB and GAMS with the Gurobi optimization solver. We start with an empty set, and then we gradually increase the number of assets adding 1 asset at a time until we find a set $K \subset \Lambda$ with $\mathrm{csupp}\,(K) \leq q$ and such that $K \underset{\mathrm{SSD}}{\succeq} \Lambda$. In each iteration, we search for the asset that increases (11) the most.

The overall procedure consists of the following steps:

For $w = 1$ to q:

1. If $w = 1$, we search for the single asset that maximizes the value of (11), thus $\mathcal{L}_{p,q}$ is a

singleton.

2. For $1 < w < q$, we solve (11) for each additional asset, and we keep the subset K with dimension w, that maximizes (11).

3. If we find a spanning set K inside the collection of all possible subsets of Λ with dimension w, then the procedure stops.

4. Else, if $w = q$, we end up with a sparse portfolio set K that "comes as close as possible" to SSD spanning its high dimensional universe of portfolios, and we evaluate the utility loss.

Given the output of the last step of the procedure above, and since in the empirical applications p is fixed, the optimal q, i.e., the one that provides the portfolio that comes closest in eliminating the empirical utility loss, can be readily estimated. To do so, and if the output of step 4 does not already imply zero optimal empirical utility loss, we may continue for $w > q$ up to p.

6 Monte Carlo

We approximate finite sample properties of our sparse SD Spanning test via two Monte Carlo (MC) experiments. We utilize data generated processes involving multivariate Gaussian distributions. Specifically, we generate T mutually independent time-series observations from N dimensional Gaussian distributions representing the joint stationary distributions of N base assets. The MC design details and results are presented below.

6.1 First experiment

The first experiment is based on a problem with $N = 49$, 100, 500 mutually i.i.d. normally distributed prospects with $T = 300$, 500, 1000 observations. Mutual iid-ness is used to invoke the aforementioned argument by Samuelson (1967) and it can be empirically motivated by the analysis of hedged returns of well-diversified portfolios. The number of prospects are selected to match the results of the empirical application, where 49, 100, and 500 assets are used.

The covariance matrix is fitted to the historical monthly returns of the three datasets used in the in-sample empirical analysis: The 49 Industry portfolios from Kenneth French's web page, the 100 assets of the FTSE 100 index, and the 500 assets of the S&P500 index. The mean vector in each case is calculated from the historical monthly returns of these assets.

Based on the empirical application, we set $q = 13$, for $N = 49$, $q = 25$, for $N = 100$, and finally, $q = 45$, for $N = 500$. For each combination of N and T, we repeat 500 times the sparse optimisation model and check how many times we get a number of assets closed to q on average. We additionally measure the variability of the loss. Table 6.1 exhibits the results of the first MC experiment.

6.2 Second experiment

The second experiment is based on a problem with $N = 50$ jointly normally distributed prospects with again $T = 300$, 500, 1000 observations. In this experiment, we evaluate the expected utility loss if q is low. We consider a set A of 5 asset returns with equal means $\mu_A = 0.3$ and equal standard deviations $\sigma_A = 0.15$, and a set B of 5 asset returns with equal means $\mu_B = 0.15$ and equal standard deviations $\sigma_B = 0.1$. Since $(\mu_A - \mu_B)/(\sigma_B - \sigma_A) < 0$, there is no portfolio in set A that dominates any portfolio in set B by SSD, and vice versa. The other 40 generated asset returns have equal means $\mu = 0.1$ and equal standard deviations $\sigma = 0.5$. The correlation coefficient of all N asset returns is set to $\rho_{i,j} = 0.001$ for any pairs of $i, j = 1, \ldots, N$, $i \neq j$. Any convex combination of assets that belong to sets A and B dominate any portfolio constructed from the other 40 assets by SSD. We set q equal to either 5 or 10. For each T, we repeat 500 times the sparse optimisation model and we measure the variability of the loss. Table 2 exhibits the results.

Sample size T	300	500	1000
Case 1: $N = 49$, $q = 13$			
Assets selected:			
Average number	11.45	12.04	12.54
St Deviation	1.18	1.12	1.13
Variability of the Loss:			
Average Loss	0.0002	0.0002	0.0001
Standard Error	0	0	0
Case 2: $N = 100$, $q = 25$			
Assets selected:			
Average number	22.57	23.02	23.88
St Deviation	1.33	1.30	1.29
Variability of the Loss:			
Average Loss	0.0001	0.0001	0.0001
Standard Error	0	0	0
Case 3: $N = 500$, $q = 45$			
Assets selected:			
Average number	42.3	42.85	43.34
St Deviation	1.68	1.57	1.54
Variability of the Loss:			
Average Loss	0.0001	0.0001	0.0001
Standard Error	0	0	0

Table 1: Monte Carlo Experiment 1. The experiment is based on a problem with $N = 49, 100, 500$ normally distributed assets and $T = 300, 500, 1000$ time series observations. We compute the average number of assets selected and the standard deviations of these. We also measure the variability of the loss, by computing the average loss and the standard error of the loss.

Sample size T	300	500	1000
Case 1: $q = 5$			
Assets selected:			
Average number	5	5	5
St Deviation	0.0	0.0	0.0
Variability of the Loss:			
Average Loss	0.01	0.009	0.008
Standard Error	0.0003	0.0003	0.0002
Case 2: $q = 10$			
Assets selected:			
Average number	10	10	10
St Deviation	0.0	0.0	0.0
Variability of the Loss:			
Average Loss	0.0	0.0	0.0
Standard Error	0.0	0.0	0.0

Table 2: Monte Carlo Experiment 2. The experiment is based on a problem with $N = 50$ normally

7 Empirical Application

In the empirical application, we analyze large data sets of equity returns to study whether sparse SSD holds or not. We investigate the performance of our strategy based on the S&P 500 index constituents, and we compare the results with the sparse mean-variance efficient portfolios of Ao, Li, and Zheng (2019). We consider the period from January 1981 to December 2020, a total of 480 monthly return observations.

7.1 In-sample Analysis

We start with the empty set, and then we add elements to it in r iterations. In each iteration, the algorithm adds to its current solution the single element decreasing the value of this solution by the most, i.e., the element with the largest marginal value with respect to the current solution. The target is to get the optimal portfolio with size q that yields the solution equal to zero (zero diversification loss). In that case, we are able to build a sparse portfolio of dimension q from a large set of assets of dimension p so that we cannot get further improvement from considering additional assets (full diversification).

In Figure 1, we observe that the number of assets that yield to zero diversification loss is 45 (upper panel)[1]. In the same Figure (lower panel), we also observe that the MAXSER portfolio of Ao, Li, and Zheng (2019) consists of 32 assets. We calculate the diversification loss, namely the estimated expected utility loss, of the optimal MAXSER portfolio with respect to the SSD portfolio with the smallest number of stocks reaching the zero bound. In the same graph, we also report the upper bound of a 95% as well as 90% one-sided confidence intervals (CI) corresponding to the portfolio reaching the zero bound (45 assets). We observe that the diversification loss of the MAXSER portfolio is between the loss of the SS-SSD portfolio for $q = 32$ and the 90% confidence interval.

[1] Analogous analysis has been done for the FTSE100 constituents as well as the 49 Industry portfolios of Kenneth French. For the FTSE100, we get a subset K with size $q = 25$ that yields zero diversification loss, while, for the 49 Industry portfolios, the size is 13 assets.

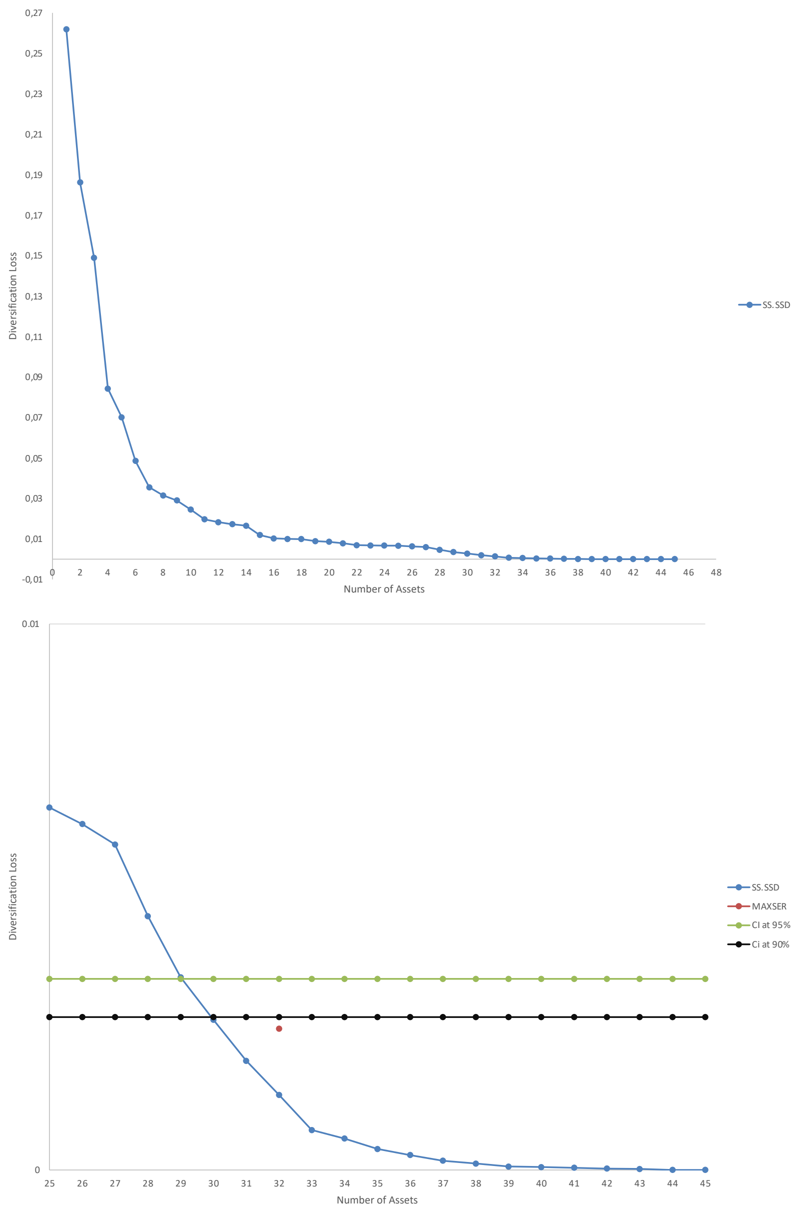

Figure 1: The upper panel plots the diversification loss with respect to the number of assets for the SS-SSD optimal portfolios. The lower panel plots the diversification loss of the optimal MAXSER portfolio with respect to the SS-SSD portfolio with zero loss, and the upper bound of a 95% and 90% one-sided confidence intervals (CI).

Table 3 reports the performance and risk measures of the in-sample performance of the MAXSER and the SS-SSD optimal portfolios as well as the $1/N$ (equally-weighted) portfolio. These measures allow us to better figure out the differences between the two portfolios. We observe that the mean as well as the standard deviation of the SS-SSD portfolio are higher that those of the MAXSER portfolio, while the Sharpe ratio is slightly lower. It is expected, since the Sharpe ratio is maximized in the construction of the MAXSER portfolio. The skewness is less negative. The VaR and the Expected Shortfall are lower as expected when investors want to mitigate the impact of large losses (left tail). The SS-SSD portfolio targets and achieves a transfer of probability mass from the left to the right tail of the return distribution when compared to the MAXSER portfolio. We also observe that both the SS-SSD and the MAXSER portfolios outperform the $1/N$ portfolio in all performance and risk measures.

Table 3: In-sample performance: risk and performance measures

Measures	MAXSER	SS-SSD	$1/N$
Average return	0.0126	0.0129	0.0133
Standard Deviation	0.0314	0.0331	0.0458
Sharpe ratio	0.4013	0.3904	0.2899
Skewness	-0.2122	-0.1986	-0.2689
Kurtosis	1.2521	1.7595	2.9690
Value-at-Risk	0.0430	0.0396	0.0615
Expected Shortfall	0.0651	0.0617	0.0959

Entries report the risk and performance measures (Sharpe ratio, Skewness, Kurtosis, VaR, ES) for the MAXSER, the SS-SSD optimal portfolios as well as the $1/N$ portfolio. The data cover the period from January, 1980 to December, 2020.

Finally, Table 4 reports the optimal average weights of the major Industries selected by each one of the two portfolios. We observe that both portfolios are well diversified and invest in almost the same Industries, with different overall weights.

Table 4: In-sample analysis: Average S&P500 Industry weights

	MAXSER	SS-SSD
Weights		
Capital Goods	4.50%	3.43%
Consumer Services	8.39%	6.57%
Financial	4.73%	3.60%
Consumer Staples	0.0%	3.24%
Food	3.21%	2.70%
Health care	7.43%	8.31%
Household	5.58%	4.37%
IMedia	4.58%	4.34%
Pharm	6.89%	5.69%
Retailing	17.21%	19.43%
Software	14.51%	16.21%
Technology	11.45%	12.79%
Transportation	5.62%	4.81%

Entries report the average Industry weights of the MAXSER and the SS-SSD portfolios in the major Industries of the S&P500 Index.

7.2 Rolling-window analysis

We conduct out-of-sample backtesting experiments and we evaluate the optimal SS-SSD portfolios achieving a zero diversification loss in a rolling-window scheme. We use a window width of 240 monthly return observations. A stock is excluded from the asset pool if it has missing data in the 240-month training period. Therefore, the number of stocks varies over time and can be smaller than the total number of constituents of the S&P500. Each month the portfolios are constructed using the monthly returns during the prior 240 months. The clock is advanced and the realized returns of the optimal portfolios are determined from the actual returns of the various assets. The same procedure is then repeated for the next time period and the ex post realized returns over the period from 01/2001 to 12/2020 (240 months) are computed. The out-of-sample test is a real-time exercise avoiding a potential look-ahead bias and mimicking the way that a real-time investor acts in practice.

We again compare the performance of the optimal SS-SSD portfolios with that of the MAXSER portfolios of Ao, Li, and Zheng (2019). The upper panel of Figure 2 plots the number of stocks of the optimal SS-SSD portfolios through time that eliminate the diversification loss, as well as the number of stocks of the efficient MAXSER portfolio. The lower panel plots the estimated expected loss of the optimal MAXSER corresponding to the inefficient SSD portfolios with the same number of stocks as MAXSER. The diversification loss is zero for the efficient SS-SSD portfolios corresponding to the upper panel by construction. On a rolling-window basis, the number of assets in the SS-SSD portfolios is always higher than in the MAXSER portfolios. It shrinks to less than 25 assets in the crisis periods of 2008-2009 and at the beginning of the Covid-19 period. Otherwise the number of assets in the SS-SSD portfolios is stable between 30 and 35. The number of assets in the MAXSER portfolios is more volatile.

Figure 3 illustrates the out-of-sample cumulative returns of the SS-SSD, the MAXSER and the $1/N$ portfolios during the period (January 2001 to December 2020). The grey areas are the NBER recession periods. We observe that the SS-SSD optimal portfolio has a 19.3 times higher value at the end of the holding period compared to the beginning, while the MAXSER portfolio has a 17.1 times higher value. The $1/N$ portfolio exhibits the worst performance, with 14.3 higher value than at the beginning of the period.

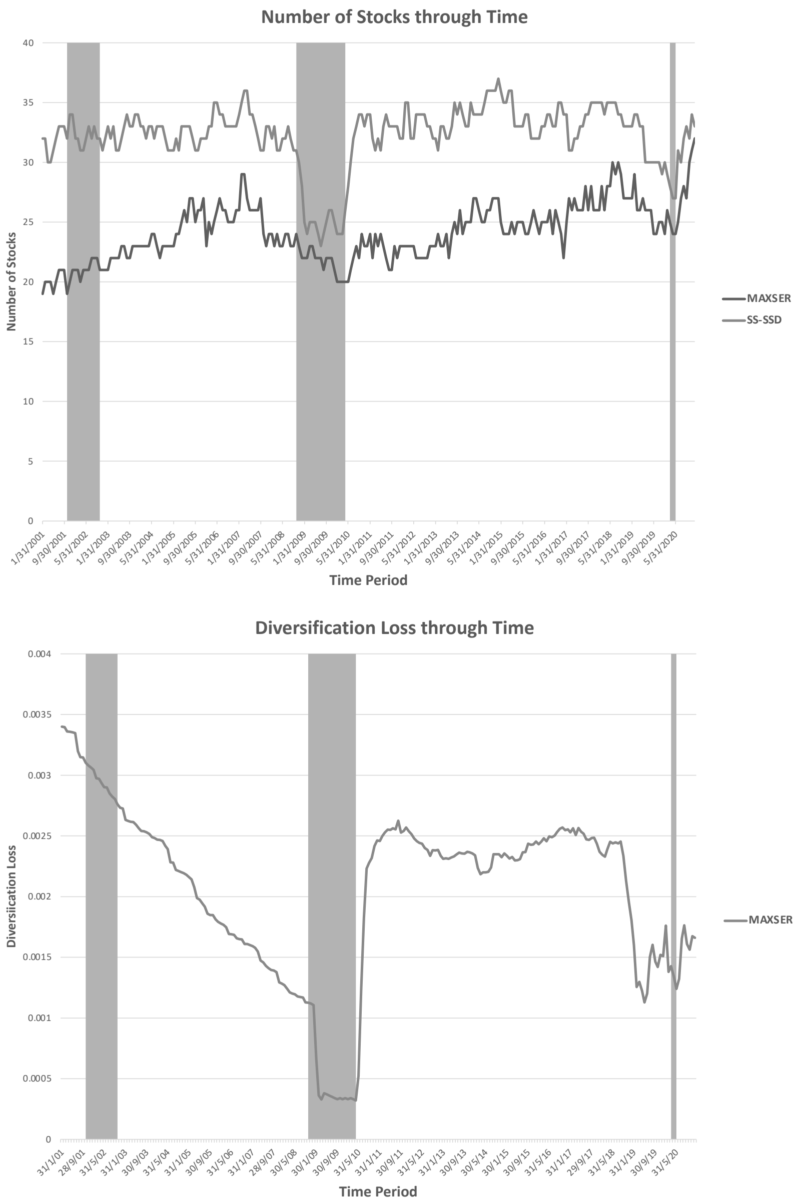

Figure 2: The upper panel plots the number of stocks of the optimal SS-SSD portfolios through time that eliminate the diversification loss, as well as the number of stocks of the efficient MV portfolios. The lower panel plots the estimated expected loss of the optimal MAXSER portfolios corresponding to the inefficient SS-SSD portfolios with the same number of stocks as MAXSER. The grey areas are the NBER recession periods

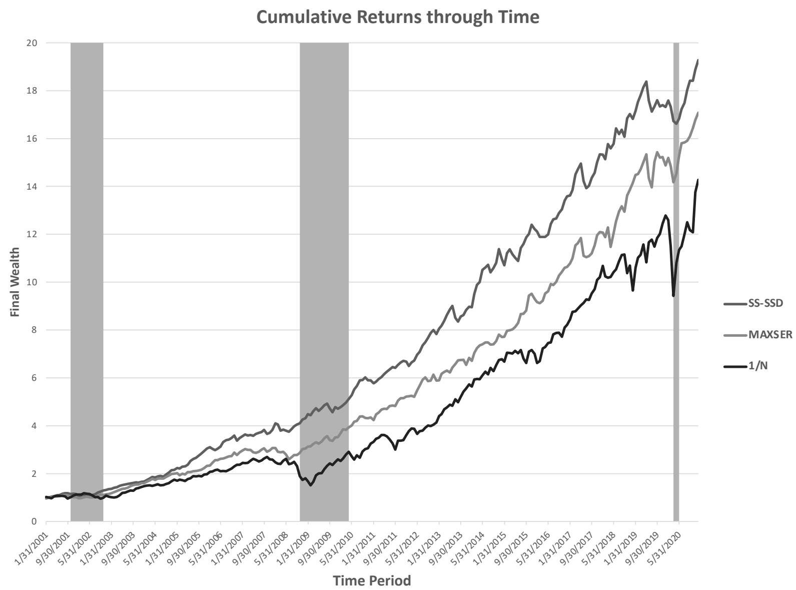

Figure 3: Cumulative performance of the MAXSER, the SS-SSD and the $1/N$ portfolios for the out-of-sample period from January 2001 to December 2020. The grey areas are the NBER recession periods.

Next, we compare the performance of the SS-SSD optimal portfolio with the performance of the MAXSER optimal portfolio using both non-parametric tests as well as parametric performance measures.

7.2.1 Non-parametric stochastic dominance performance test

We use the pairwise (non-)dominance test of Anyfantaki et al. (2022), for a risk-adjusted comparison of the out-of-sample performance of the SS-SSD and MAXSER portfolios.

The definition for second order stochastic non-dominance is the following:

Definition 4. (Stochastic non-dominance): *The SS-SSD portfolio λ does not strictly second order stochastically dominate the MAXSER portfolio κ*, say $\lambda \not\succ_F \kappa$, iff

$$\exists z \in \mathcal{Z} : \, D(z, \lambda, \kappa, F) > 0, \text{ or } \forall z \in \mathcal{Z} : \, D(z, \lambda, \kappa, F) = 0.$$

Strict second order stochastic non-dominance holds iff κ achieves a higher expected utility for some non-decreasing and concave utility function or achieves the same expected utility for every non-decreasing and concave utility function. Equivalently, strict stochastic non-dominance holds iff κ is strictly preferred to λ by some risk averter, or every risk averter is indifferent between them.

We test the null hypothesis H_0' vis-à-vis the alternative hypothesis :

H_0': SS-SSD portfolio λ does not strictly second order stochastically dominate MAXSER portfolio κ,

H_1': SS-SSD Portfolio λ stochastically dominates MAXSER portfolio κ,

For the pairwise test of the two portfolios, the test statistic is the following:

$$\xi_T = \sup_{z \in \mathcal{Z}} D\left(z, \kappa, \lambda, F\right). \tag{13}$$

To calculate the p-value, we use block-boostrapping. The p-value is approximated by $\tilde{p}_j = \frac{1}{R}\sum_{r=1}^{R}\{\xi_{T,r}^{\star} > \xi_T\}$, where ξ_T is the test statistic, $\xi_{T,r}^{\star}$ is the bootstrap test statistic, averaging over $R = 1000$ replications. We reject the null hypothesis of non-dominance if the p-value is lower than 5%.

The test statistic ξ_T is -0.0012, and the p-value is estimated at 4.4%. Thus, we reject the null hypothesis of non-dominance of portfolio SS-SSD over MAXSER.

7.2.2 Performance summary of the optimal portfolios

We also compute a number of commonly used parametric performance measures for portfolios: the Sharpe ratio, the downside Sharpe ratio of Ziemba (2005), the 95% Value-at-Risk (with a positive sign for a loss), the 95% Expected Shortfall (with a positive sign for a loss), the upside potential and downside risk (UP) ratio of Sortino and van den Meer (1991), the portfolio turnover, and a measure of the portfolio risk-adjusted returns net of transaction costs (DeMiguel et al. (2009)) and the opportunity cost.

The definition of downside Sharpe ratio uses the downside variance (or more precisely the downside risk) defined as

$$\sigma^2_{P_-} = \frac{\sum_{t=1}^{T} (R_t)^2_-}{T-1}, \tag{14}$$

where R_t is the return of portfolio P at day t which is below zero (i.e., those with losses).

Given that the total variance equals twice the downside variance $2\sigma^2_{P_-}$, the downside Sharpe ratio is given by

$$S_P = \frac{\bar{R}_P - \bar{R}_f}{\sqrt{2}\sigma_{P-}}, \tag{15}$$

where $\bar{R}_P$ is the average period return of portfolio P and $\bar{R}_f$ is the average risk free rate.

The UP ratio compares the upside potential to the shortfall risk over a specific target (benchmark):

$$\text{UP ratio} = \frac{\frac{1}{T}\sum_{t=1}^{T} (R_{P,t} - R_{f,t})_+}{\sqrt{\frac{1}{T}\sum_{t=1}^{T} ((R_{f,t} - R_{P,t})_+)^2}}, \tag{16}$$

where $R_{P,t}$ is the realized monthly return of the portfolio P for the out-of-sample period, T is the number of experiments performed, and $R_{f,t}$ is the monthly return of the benchmark (the riskless asset). The numerator equals the average excess return over the benchmark reflecting the upside potential while the denominator measures the downside risk (i.e., shortfall risk over the benchmark).

Both the downside Sharpe and UP ratios are viewed to be more appropriate measures of performance than the typical Sharpe ratio given the asymmetric return distribution of assets.

The portfolio turnover (PT) measures the degree of rebalancing required to implement each one of the two strategies. For any portfolio strategy P, the portfolio turnover is defined as the average of the absolute change of weights over the T rebalancing points in time and across the N available assets:

$$PT = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N} (|w_{P,i,t+1} - w_{P,i,t}|), \tag{17}$$

where $w_{P,i,t+1}$ and $w_{P,i,t}$ are the optimal weights of asset i under strategy P (SS-SSD or MAXSER) at time t and $t+1$, respectively.

The performance of the portfolios is also assessed under the risk-adjusted (net of transaction costs) returns measure of DeMiguel et al. (2009) which is an indicator of how the proportional transaction cost generated by the portfolio turnover affects the portfolio returns. We use a transaction cost of 35 bps, which is typical in the literature. For this, the change in the net of transaction cost wealth NW_P of portfolio P through time is first defined as

$$NW_{P,t+1} = NW_{P,t}(1 + R_{P,t+1})[1 - trc \times \sum_{i=1}^{N}(|w_{P,i,t+1} - w_{P,i,t}|), \quad (18)$$

where trc is the proportional transaction cost and $R_{P,t+1}$ is the realized return of portfolio P at time $t+1$.

Then, the portfolio return net of transaction costs is defined as

$$RTC_{P,t+1} = \frac{NW_{P,t+1}}{NW_{P,t}} - 1. \quad (19)$$

The return-loss measures the additional return needed so that the MAXSER optimal portfolio performs equally well with the SS-SSD portfolio is defined as

$$R_{Loss} = \frac{\mu_{SSD}}{\sigma_{SSD}} \times \sigma_{MAXSER} - \mu_{MAXSER}, \quad (20)$$

where μ_{MAXSER} and μ_{SSD} are the out-of-sample mean of monthly RTC for the MAXSER and the SS-SSD opportunity set respectively, and σ_{MAXSER} and σ_{SSD} are the corresponding standard deviations.

Finally, the opportunity cost θ of Simaan (2013) is used, which is a useful measure for the economic significance of the performance difference of two portfolios. It is defined as the return that needs to be added to (or subtracted from) the MAXSER portfolio return R_{MAXSER}, so that the investor is indifferent (in utility terms) between the the two different portfolios.

$$E[U(1 + R_{MAXSER} + \theta)] = E[U(1 + R_{SS-SSD})]. \tag{21}$$

A positive (negative) opportunity cost implies that the investor is better (worse) off if he invests in the SS-SSD over the MAXSER portfolio.

Given that this measure takes into account the entire probability density function of asset returns, it is suitable to evaluate strategies even when the asset return distribution is not normal. For the calculation of the opportunity cost, exponential and power utility functions are used, consistent with second degree stochastic dominance. For the coefficient of risk aversion, alternative values are employed.

Table 5 reports the parametric performance measures for the MAXSER, the SS-SSD optimal portfolios and the $1/N$ portfolio for the sample period. The higher the value of each one of these measures, the greater the investment opportunities for the relative portfolio.

We observe that the average returns, the Sharpe ratios and the downside Sharpe ratios of the SS-SSD optimal portfolios are higher than those of the MAXSER optimal portfolios. It reflects an increase in the risk-adjusted performance (i.e., an increase in the expected return per unit of risk) and hence expands the investment opportunities for risk-averse investors. The same is true for the UP ratio. The Value-at-Risk and the Expected Shortfall (with a positive sign for a loss) of the SS-SSD portfolios are lower, indicating lower downside losses. Furthermore, the MAXSER portfolios induce slightly less portfolio turnover than the SS-SSD portfolios. The return-loss measure that takes into account transaction costs, is positive. Finally, the opportunity cost is always positive, indicating that a positive return should be added in the MAXSER portfolio to achieve the same expected return with the SS-SSD portfolio. The $1/N$ portfolio exhibits again the worst performance, dominated by both the SS-SSD as well as the MAXSER portfolios.

Table 5: Out-of-sample performance: risk and performancemeasures

Measures	MAXSER	SS-SSD	$1/N$
Average return	0.0122	0.0127	0.0121
Standard Deviation	0.0258	0.0239	0.0450
Sharpe ratio	0.4056	0.4571	0.2313
Downside Sharpe Ratio	0.8614	1.1188	0.9311
Value-at-Risk	0.0403	0.0295	0.0744
Expected Shortfall	0.0532	0.0476	0.1004
UP ratio	1.0864	1.2014	0.7704
Portfolio Turnover	8.477%	8.835%	0.0
Return Loss	0.087%		0.156%
Opportunity Cost			
Exponential Utility			
ARA=2	0.073%		0.126%
ARA=4	0.081%		0.139%
ARA=6	0.092%		0.152%
Power Utility			
RRA=2	0.070%		0.132%
RRA=4	0.079%		0.144%
RRA=6	0.091%		0.159%

Entries report the risk and performance measures (Sharpe ratio, Downside Sharpe ratio, VaR, ES, UP ratio, Portfolio Turnover, Returns Loss and opportunity cost) for the MAXSER, the SS-SSD and the $1/N$ portfolios. The realized monhtly returns cover the period from January, 2001 to December, 2020.

Let us now analyze the composition of the SS-SSD and the MAXSER portfolios through time. Figure 4 reports the optimal average weights of the major Industries selected by each one of the two portfolios during the out-of-sample period. We observe that both portfolios are well diversified and invest in almost the same Industries, with different overall weights. The optimal SS-SSD portfolio invests mainly in 9 industry sectors with a larger weighting on small size, high book-to-market, and momentum stocks from the S&P 500 index.

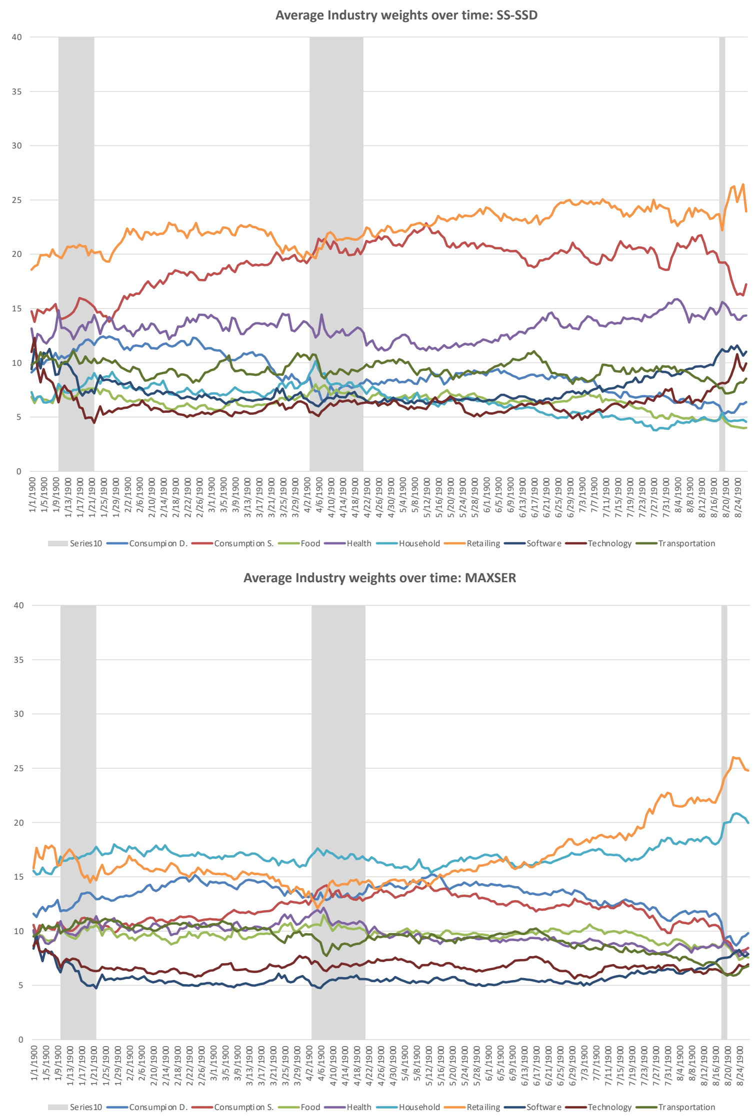

Figure 4: Average Industry weights through time. The upper panel plots the average Industry weights of the optimal SS-SSD portfolios, while the lower panel plots the average Industry weights of the optimal MAXSER portfolios, for the out-of-sample period from January 2001 to December 2020. The grey areas are the NBER recession periods.

Figures 5 and 6 exhibit the range of the Alpha and Beta coefficients of the individual stocks of these portfolios during the out-of-sample period. For the calculations of the Alpha and Beta coefficients, the previous 5 years of individual monthly returns have been used (60 monthly returns). We can observe that the Beta coefficients in the SS-SSD portfolios have a more defensive profile. The heterogeneity of Alpha and Beta coefficients is lower in the SS-SSD portfolios.

Figure 5: The upper panel plots the range of the Alpha coefficients of the individual stocks of the optimal SS-SSD portfolios through time. The lower panel plots the range of the Alpha coefficients of the individual stocks of the optimal SS-SSD portfolios through time. The grey areas are the NBER recession periods.

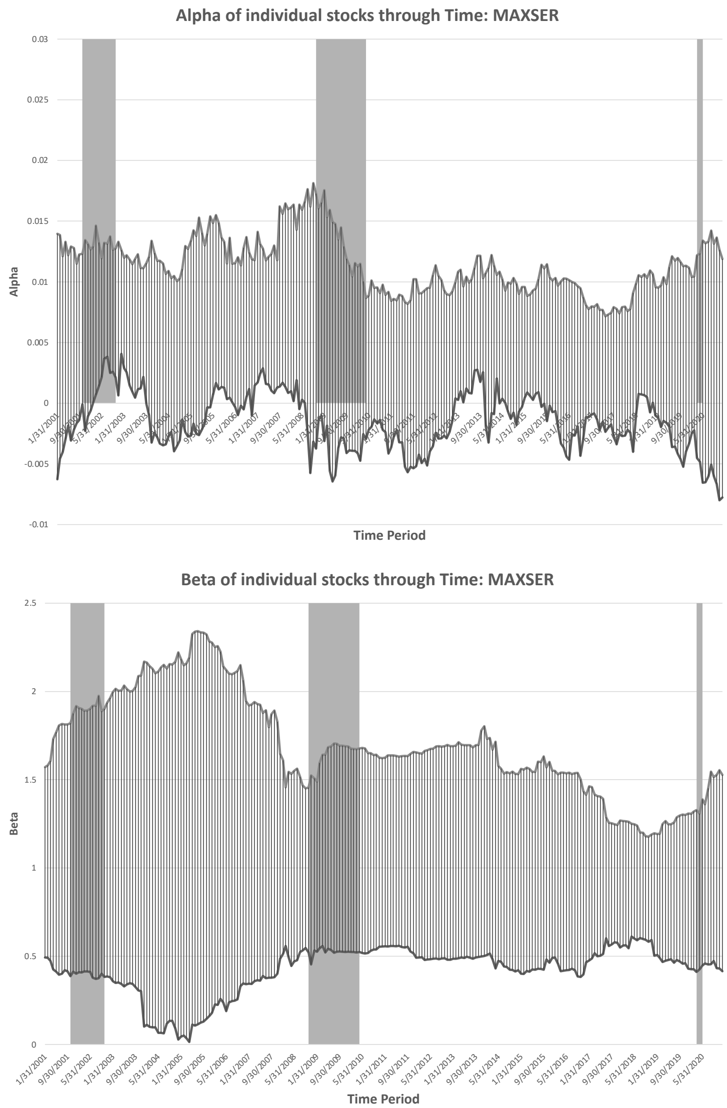

Figure 6: The upper panel plots the range of the Alpha coefficients of the individual stocks of the optimal MAXSER portfolios, and the lower panel plots the range of the Beta coefficients of the individual stocks of the optimal MAXSER portfolios, for the out-of-sample period from January 2001 to December 2020. The grey areas are the NBER recession periods.

Finally, we investigate which factors explain the returns of the active investors with SSD preferences. To do so, we start with the classical single factor model, and we additionally use five asset pricing models that are popular in the literature. Namely, we use, the Fama-French 6-factor model (2016), the q-factor model of Hou, Xue and Zhang, (2015), the M4 factor model of Stambaugh and Yuan, (2017), the Barillas and Shanken 6-factor model (2018) and the 3-factor model of Daniel, Hirshleifer, and Sun (2020). The last is included to give an economic insight on behavioral influence.

First, we consider linear regression models of the following form:

$$R_{p,t} - R_{f,t} = a_i + \sum_i b_i R_{i,t} + e_{i,t},$$

where $R_{p,t}$ is the return of either the MAXSER or SS-SSD optimal portfolio at period t, and, $R_{i,t}$ is the return on the i_{th} factor. If the exposures b_i to the various factors capture all variation in expected returns, the intercept a_i is zero since the factors are tradable.

	a_i	$R_M - R_F$
MaxSer		
Coef.	0.0105	-0.0148
t-stat	5.772	-0.348
p-values	0.0	0.7275
SS-SSD		
Coef.	0.0119	-0.0077
t-stat	7.186	-0.201
p-values	0.0	0.8411

Table 6: Single factor model (CAPM). Entries report the coefficients and their respective t-statistics and p-values for the MAXSER portfolio (upper panel) and the SS-SSD portfolio (lower panel). The dataset spans 01/2001-12/2020 for optimal portfolios computed with 240-month windows rolled over one month.

	a_i	$R_M - R_F$	$PEAD$	FIN
MaxSer				
Coef.	0.0095	0.0750	0.0219	0.1231
t-stat	4.7770	1.3773	0.2244	2.0896
p-values	0.0	0.1700	0.8227	0.0380
SS-SSD				
Coef.	0.0124	0.0293	-0.0738	0.0696
t-stat	6.5221	0.5643	-0.7938	1.2375
p-values	0.0	0.5732	0.4283	0.2174

Table 7: Daniel, Hirshleifer, and Sun (2020), three-factor model. Entries report the coefficients and their respective t-statistics and p-values for the MAXSER portfolio (upper panel) and the SS-SSD portfolio (lower panel). The dataset spans 01/2001-12/2016 for optimal portfolios computed with 240-month windows rolled over one month.

	a_i	$R_M - R_F$	SMB	$R - IA$	$R - ROE$	$HML - AQR$	$UMD - AQR$
MaxSer							
Coef.	0.0104	0.0004	-0.0695	0.0492	0.0123	0.0466	0.0148
t-stat	5.4458	0.0079	-0.8498	0.3756	0.1071	0.4898	0.2104
p-values	0.0	0.9937	0.3964	0.7075	0.9148	0.6248	0.8336
SS-SSD							
Coef.	0.0116	0.0065	-0.0808	-0.0634	0.0666	0.1202	0.0261
t-stat	6.7464	0.1350	-1.0955	-0.5364	0.6409	1.4021	0.4103
p-values	0.0	0.8927	0.2745	0.5922	0.5222	0.1623	0.6820

Table 8: Barillas and Shanken (2018), six factor model. Entries report the coefficients and their respective t-statistics and p-values for the MAXSER portfolio (upper panel) and the SS-SSD portfolio (lower panel). The dataset spans 01/2001-12/2020 for optimal portfolios computed with 240-month windows rolled over one month.

	a_i	$R_M - R_F$	SMB	HML	RMW	CMA	Mom
MaxSer							
Coef.	0.0104	0.0044	-0.0586	-0.0472	0.0180	0.1355	-0.0154
t-stat	5.3465	0.0795	-0.7455	-0.5413	0.1605	1.0840	-0.3541
p-values	0.0	0.9367	0.4567	0.5888	0.8726	0.2796	0.7236
SS-SSD							
Coef.	0.0121	-0.0040	-0.0708	0.0770	0.0153	-0.0503	-0.0139
t-stat	6.8617	-0.0788	-0.9930	0.9747	0.1512	-0.4443	-0.3544
p-values	0.0	0.9372	0.3218	0.3308	0.8799	0.6572	0.7234

Table 9: Fama-French (2016), six factor model. Entries report the coefficients and their respective t-statistics and p-values for the MAXSER portfolio (upper panel) and the SS-SSD portfolio (lower panel). The dataset spans 01/2001-12/2020 for optimal portfolios computed with 240-month windows rolled over one month.

	a_i	$R_M - R_F$	$SMB1$	$MGMT1$	$PERF1$
MaxSer					
Coef.	0.0105	0.0045	-0.0131	0.1789	-0.0710
t-stat	5.2433	0.0777	-0.1626	2.4328	-1.4994
p-values	0.0	0.9381	0.8710	0.0159	0.1355
SS-SSD					
Coef.	0.0129	0.0026	-0.0498	0.1094	-0.0520
t-stat	6.7428	0.0464	-0.6478	1.5526	-1.1453
p-values	0.0	0.9630	0.5179	0.1222	0.2535

Table 10: Stambaugh and Yuan(2017), M4 four-factor model. Entries report the coefficients and their respective t-statistics and p-values for the MAXSER portfolio (upper panel) and the SS-SSD portfolio (lower panel). The dataset spans 01/2001-12/2016 for optimal portfolios computed with 240-month windows rolled over one month.

	a_i	$R_M - R_F$	ME	IA	ROE
MaxSer					
Coef.	0.0106	0.0301	-0.0521	0.1102	0.0181
t-stat	5.5931	0.5541	-0.6879	1.0715	0.2213
p-values	0.0	0.5801	0.4923	0.2852	0.8251
SS-SSD					
Coef.	0.0122	0.0263	-0.0452	0.0503	0.0331
t-stat	6.9438	0.5218	-0.6424	0.5270	0.4361
p-values	0.0	0.6024	0.5213	0.5987	0.6632

Table 11: Hou, Xue and Zhang, (2015) q- four-factor model. Entries report the coefficients and their respective t-statistics and p-values for the MAXSER portfolio (upper panel) and the SS-SSD portfolio (lower panel). The dataset spans 01/2001-12/2020 for optimal portfolios computed with 240-month windows rolled over one month.

Tables 7- 11 reports the coefficient estimates of the factor models, as well as their respective t-statistics and p-values. The results indicate that none of the factor models could fully explain the performance of the two strategies. In particular, a close to zero market loading indicates a market neutral exposure. The intercept a_i is statistically different from zero in all cases. We also observe that the only factors that are significant for the MAXSER returns are the FIN factor of the 3-factor model of Daniel, Hirshleifer, and Sun (2020), and the MGMT1 factor of the Stambaugh and Yuan(2017), four-factor model. On the other hand, there is no statistically significant factor that explains the returns of the SS-SSD portfolios. The results indicate that perhaps other factors drive the performance of the these portfolios.

In both factor models, we observe that the beta market is slightly smaller than one (defensive) for both portfolios as expected. The negative sign for the SMB factor loading and positive sign for the HML factor loading correspond to an additional defensive tilt of the SS-SSD portfolio returns. Defensive strategies overweight large value stocks and underweight small growth stocks (Novy-Marx (2016)).

8 Concluding remarks

Our new methodology designed to target sparse spanning portfolios shows that we can often limit ourselves to a subset of a large investment opportunity set without sacrifying expected utility because of under-diversification. It also reveals that a sparse mean-variance portfolio selection yields under-diversification w.r.t. an optimal sparse spanning portfolio. This paper focuses on second-order stochastic dominance but could be modified to accommodate higher-order stochastic dominance. We could then check whether the empirical findings extend in such settings as well.

Acknowledgements: We thank J.-C. Rochet, participants at SFI Research days and FEM 2023, and participants at seminars at ....

References

[1] Aliprantis C. D., and K. C. Border, 2006. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer.

[2] Anderson, A., 2013. Trading and Under-Diversification. Review of Finance 17, 1699-1741.

[3] Andrews, D. W., 1994. Empirical Process Methods in Econometrics, in *Handbook of Econometrics*, Vol. 4, Edited by R.F. Engle and D.L. McFadden, 2247-2294.

[4] Anyfantaki, S., Maasoumi, E., Ren, J. and N. Topaloglou, 2022. Evidence of Uniform Inefficiency in Market Portfolios based on Dominance Tests. Journal of Business and Economic Statistics 40(3), 937-949.

[5] Ao, M., Li, Y., and Zheng, X., 2019. Approaching Mean-Variance Efficiency for Large Portfolios. Review of Financial Studies 32(7), 2890-2919.

[6] Arvanitis, S., Hallam, M. S., Post, T., and Topaloglou, N., 2019. Stochastic Spanning. Journal of Business and Economic Statistics 37(4), 573-585.

[7] Arvanitis, S., Post, T., and Topaloglou, N., 2021. Stochastic Bounds for Reference Sets in Portfolio Analysis. Management Science 67(12), 7737-7754.

[8] Arvanitis, S., Scaillet, O., and Topaloglou, N., 2020a. Spanning Analysis of Stock Market Anomalies under Prospect Stochastic Dominance. Forthcoming in Management Science.

[9] Arvanitis, S., Scaillet, O., and Topaloglou, N., 2020b. Spanning Tests for Markowitz Stochastic Dominance. Journal of Econometrics 217(2), 291-311.

[10] Arvanitis, S., and Topaloglou, N., 2017. Testing for Prospect and Markowitz Stochastic Dominance Efficiency. Journal of Econometrics 198(2), 253-270.

[11] Athayde G., and Flores, R., 2004. Finding a Maximum Skewness Portfolio: A General Solution to Three-Moments Portfolio Choice. Journal of Economic Dynamics and Control 28, 1335-1352.

[12] Barillas, F. and Shanken, J. (2016). Which Alpha? Review of Financial Studies, 30(4), 1316-1338.

[13] Beer, G., and Lucchetti, R., 1991. Convex Optimization and the Epi-Distance Topology. Transactions of the American Mathematical Society 327(2), 795-813.

[14] Berk, J., 1997. Necessary Conditions for the CAPM. Journal of Economic Theory 73, 245-257.

[15] Bertsimas, D., and Tsitsiklis, J., 1997. *Introduction to Linear Optimization*. 1st edition, Athena Scientific, Belmont.

[16] Bian, A., Buhmann, J., Krause, A., and Tschiatschek, S., 2017. Guarantees for Greedy Maximization of Non-submodular Functions with Applications. In Proceedings of the 34th International Conference on Machine Learning, 498-507.

[17] Borwein, J., and A.S. Lewis, 2010. *Convex Analysis and Nonlinear Optimization: Theory and Examples*. Springer Science and Business Media.

[18] Brodie, J., Daubechies, I., DeMol, C., Giannone, D., and Loris, I., 2009. Sparse and Stable Markowitz Portfolios. Proceedings of the National Academy of Sciences 106(30), 12267-12272.

[19] Buchbinder, N., and Feldman, M., 2018. Submodular Functions Maximization Problems - A Survey in *Handbook of Approximation Algorithms and Metaheuristics*, 2nd Ed., Ed. T. Gonzalez, Vol. 1, Chapman and Hall, London, 753-788.

[20] Caner, M., Medeiros, M., and Vasconcelos, G., 2021. Residual Based Nodewise Regression in Factor Models with Ultra-High Dimensions: Analysis of Mean-Variance Portfolio Effiiency and Estimation of Out-of-Sample and Constrained Maximum Sharpe Ratios. PUC working paper.

[21] Campbell, J., 2006. Household Finance. Journal of Finance 61(4), 1553-1604.

[22] Calvet, L., Campbell, J., and Sodini, P., 2007. Down or Out: Assessing the Welfare Costs of Household Investment Mistakes. Journal of Political Economy 115(5), 707-747.

[23] Chamberlain, G., 1983. A Characterization of the Distributions that Imply Mean-Variance Utility Functions. Journal of Economic Theory 29, 185-201.

[24] Cong, Y., Chen, B., and Zhou, M., 2017. Fast Simulation of Hyperplane-Truncated Multivariate Normal Distributions. Bayesian Analysis 12(4), 1017-1037.

[25] Daniel, K., Hirshleifer, D., and L. Sun, 2020, Short- and Long-hHrizon Behavioral Factors, Review of Financial Studies 33(4), 1673-1736.

[26] Danthine, J.P., and Donaldson, J., 2014. *Intermediate Financial Theory*. 3rd edition, North-Holland, Amsterdam.

[27] Das, A. and Kempe, D., 2011. Submodular Meets Spectral: Greedy Algorithms for Subset Selection, Sparse Approximation and Dictionary Selection. In Proceedings of the 28th International Conference on Machine Learning, 1057-1064.

[28] Davidson, J., 1994. *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press.

[29] DeMiguel, V., Garlappi, L., Nogales, F. J., and Uppal, R., 2009. A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms. Management Science 55(5), 798-812.

[30] DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus Naive Diversification: How Inefficient is the 1/n portfolio Strategy? Review of Financial Studies, 22, 1915-1953.

[31] De Roon, F. , Nijman, T., and Werker, B., 2001. Testing for Mean-Variance Spanning with Short Sales Constraints and Transaction Costs: the Case of Emerging Markets. Journal of Finance 56, 721-742.

[32] Dimmock, S. G., Kouwenberg, R., Mitchell, O. S., and Peijnenburg, K., 2021. Household Portfolio Underdiversification and Probability Weighting: Evidence from the Field. Review of Financial Studies 34(9), 4524-4563.

[33] Doukhan, P., Massart, P., and Rio, E., 1995. Invariance Principles for Absolutely Regular Empirical Processes. Annales de l'IHP Probabilités et Statistiques 31(2), 393-427.

[34] Edelsbrunner, H., 2014. *A Short Course in Computational Geometry and Topology* (No. Mathematical Methods). Berlin, Germany: Springer.

[35] Elenberg, E. R., Khanna, R., Dimakis, A. G., and Negahban, S., 2018. Restricted Strong Convexity Implies Weak Submodularity. Annals of Statistics 46(68), 3539-3568.

[36] Estrada, R., and Kanwal, R. P., 2012. *A Distributional Approach to Asymptotics: Theory and Applications*. Springer Science and Business Media.

[37] Elton, E., and Gruber, M., 1977. Risk Reduction and Portfolio Size: An Analytical Solution. Journal of Business 50(4), 415-437.

[38] Evans, J., and Archer, S., 1968. Diversification and the Reduction of Dispersion-an Empirical Analysis. Journal of Finance 23(5), 761-767.

[39] Fama, E. F. and French, K. R. 2016. Dissecting Anomalies with a Five-Factor Model. Review of Financial Studies, 29(1), 69-103.

[40] Fan, J., Liao, Y., and Shi, X., 2015. Risks of Large Portfolios. Journal of Econometrics 186(2), 367-387.

[41] Fan, J., Zhang, J., and Yu, K., 2012. Vast Portfolio Selection with Gross-Exposure Constraints. Journal of the American Statistical Association 107(498), 592-606.

[42] Fang, Z., and Santos, A., 2014. Inference on Directionally Differentiable Functions. arXiv preprint arXiv:1404.3763.

[43] Francq, C., and Zakoian, J.M., 2011. *GARCH Models: Structure, Statistical Inference and Financial Applications*. John Wiley and Sons, New York.

[44] Giné, E., and Nickl, R., 2016. *Mathematical Foundations of Infinite-Dimensional Statistical Models (Vol. 40)*. Cambridge University Press.

[45] Hadar, J., and Russell, W., 1969. Rules for Ordering Uncertain Prospects. American Economic Review 59, 2-34.

[46] Hanoch, G., and Levy, H., 1969. The Efficiency Analysis of Choices Involving Risk. Review of Economic Studies 36, 335-346.

[47] Hiriart-Urruty, J. B., and Lemaréchal, C., 2004. *Fundamentals of Convex Analysis*, Springer.

[48] Hiriart-Urruty, J. B., and Lemaréchal, C., 2013. *Convex Analysis and Minimization Algorithms I: Fundamentals (Vol. 305)*. Springer Science and Business media.

[49] Hong, H., and Scaillet, O., 2006, A Fast Subsampling Method for Nonlinear Dynamic Models. Journal of Econometrics 133, 557-578.

[50] Hou, K., Chen, X., and Zhang, L., 2015, Digesting Anomalies, Review of Financial Studies 28, 650-705.

[51] Huberman, G., and Kandel, S., 1987. Mean-Variance Spanning. Journal of Finance 42, 873-888.

[52] Jagannathan, R., and Ma, T., 2003. Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. Journal of Finance 58(4), 1651-84.

[53] Jobson, J., and Korkie, B., 1989. A Performance Interpretation of Multivariate Tests of Asset Set Intersection, Spanning and Mean-Variance Efficiency. Journal of Financial and Quantitative Analysis 24, 185-204.

[54] Kan, R., and Zhou, G., 2012. Tests of Mean-Variance Spanning. Annals of Economics and Finance 13(1), 145-193.

[55] Kim, J. and Pollard, D., 1990. Cube root asymptotics. Annals of Statistics 18(1),191-219.

[56] Klein, E., and A. Thompson, 1984. *Theory of Correspondences*. Wiley-Interscience, New York.

[57] Klemkosky, R., and Martin, J., 1975. The Effect of Market Risk on Portfolio Diversification. Journal of Finance 30(1), 147-154.

[58] Koijen, R., and Yogo, M., 2019. A Demand System Approach to Asset Pricing. Journal of Political Economy 127(4), 1475-1515.

[59] Kuosmanen, T., 2004. Efficient Diversification According to Stochastic Dominance Criteria. Management Science 50, 1390-1406.

[60] Ledoit, O., and Wolf, M., 2017. Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks. Review of Financial Studies 30, 4349-88.

[61] Linton, O., Maasoumi, E., and Whang, Y.J., 2005. Consistent Testing for Stochastic Dominance under General Sampling Schemes. Review of Economic Studies, 72, 735-765.

[62] Linton, O., Post, T., and Whang, Y.J., 2014. Testing for the Stochastic Dominance Efficiency of a Given Portfolio. The Econometrics Journal, 17(2), 59-74.

[63] Liu, H., 2014. Solvency Constraint, Underdiversification, and Idiosyncratic Risks. Journal of Financial and Quantitative Analysis 49(2), 409-430.

[64] Markowitz, H., 1952. Portfolio Selection. Journal of Finance 7(1), 77-91.

[65] Meinshausen, N., and Buehlmann, P., 2006. Consistent Neighbourhood Selection for High-dimensional Graphs with the Lasso. Annals of Statistics 34(3), 2006.

[66] Mencia, J., and Sentana, E., 2009. Multivariate Location-Scale Mixtures of Normals and Mean-Variance-Skewness Portfolio Allocation. Journal of Econometrics 153, 105-121.

[67] Molchanov, I., 2006. *Theory of Random Sets*. Springer.

[68] Nemhauser, G., and Wolsey, L., 1978. Best Algorithms for Approximating the Maximum of a Submodular Set Function. Mathematics of Operations Research 3(3), 177-188.

[69] Nemhauser, G., Wolsey, L., and Fisher, M., 1978. An Analysis of Approximations for Maximizing Submodular Set Functions - I. Mathematical Programming 14, 265-294.

[70] Nesterov, Y., 1998, *Introductory Lectures on Convex Optimization: A Basic Course*. Springer.

[71] Nickl, R., and Pötscher, B. M., 2007. Bracketing Metric Entropy Rates and Empirical Central Limit Theorems for Function Classes of Besov-and Sobolev-type. Journal of Theoretical Probability, 20(2), 177-199.

[72] Novy-Marx, R., 2013, The Other Side of Value: The Gross Profitability Premium, Journal of Financial Economics 108 (1), 1-28.

[73] Owen, J., and Rabinovitch, R., 1983. On the Class of Elliptical Distributions and their Applications to the Theory of Portfolio Choice. Journal of Finance 38, 745-752.

[74] Politis, D. N., Romano, J. P., and Wolf, M., 1999. *Subsampling*. Springer Science and Business Media.

[75] Post, T., 2003. Empirical Tests for Stochastic Dominance Efficiency. Journal of Finance 58, 1905-1932.

[76] Penaranda, F., and Sentana, E., 2012. Spanning Tests in Portfolio and Stochastic Discount Factor Mean-Variance Frontiers: A Unifying Approach. Journal of Econometrics 170, 303-324.

[77] Radulović, D. (1998). On the Subsample Bootstrap Variance Estimation. Test 7(2), 295-306.

[78] Rothschild, M., and Stiglitz, J., 1970. Increasing Risk: I. A Definition. Journal of Economic Theory 2(3), 225-243.

[79] Rio, E., 2017. *Asymptotic Theory of Weakly Dependent Random Processes (Vol. 80)*, Berlin: Springer.

[80] Russell, W.R. and Seo, T.K., 1989. *Representative Sets for Stochastic Dominance Rules*, in *Studies in the Economics of Uncertainty*: In honor of Josef Hadar, Springer-Verlag.

[81] Samorodnitsky, G., 1991. Probability Tails of Gaussian Extrema. Stochastic Processes and their Applications 38(1), 55-84.

[82] Samuelson P. A., 1967, General proof that diversification pays. Journal of Financial and Quantitative Analysis 2(1),1-13.

[83] Savare, G., 1996. On the Regularity of the Positive Part of Functions. Nonlinear Analysis, Theory, Methods and Applications 27(9), 1055-1074.

[84] Simaan, Y., 1993. Portfolio Selection and Asset Pricing-Three-Parameter Framework. Management Science 39, 568-577.

[85] Sion, M., 1958. On General Minimax Theorems. Pacific Journal of Mathematics 8(1), 171-176.

[86] Sortino, F. and van der Meer, R. 1991. Downside Risk-Capturing What's at Stake in Investment Situations. Journal of Portfolio Management, 17, 27-31.

[87] Stambaugh, R.F., and Yuan, Y., 2016, Mispricing Factors, Review of Financial Studies 30, 1270-1315.

[88] Trench, W.F., 1999. Asymptotic Distribution of the Spectra of a Class of Generalized Kac-Murdock-Szego Matrices. Linear Algebra and its Applications, 294, 181-192.

[89] Tibshirani, R., 1996. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society: Series B 58 (1), 267-288.

[90] van der Vaart, A.W., 2000. *Asymptotic Statistics*. Cambridge University Press.

[91] van der Vaart, A. W., and Wellner, J. A., 1996. *Weak Convergence and Empirical Processes*. Springer, New York, NY.

[92] Van Nieuwerburgh, S., and Weldkamp, L., 2010. Information Acquisition and Under-diversification. Review of Economic Studies 77, 779-805.

[93] Wainwright, M. J., 2019. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint (Vol. 48)*. Cambridge University Press.

[94] Zhang, T., 2009. On the Consistency of Feature Selection using Greedy Least Squares Regression. Journal of Machine Learning Research 10, 555-568.

[95] Zhao, P., and Yiu, B., 2006. On Model Selection Consistency of Lasso. Journal of Machine Learning Research 7, 2541-2563.

[96] Ziemba, W., 2005. The Symmetric Downside Risk Sharpe Ratio. Journal of Portfolio Management, 32(1), 108-122.

Appendix

The appendix contains the proofs of our results.

Proof of Lemma 1. The result is obtained by exploiting the continuity of D w.r.t. its first triple of arguments, and the compactness of the parameter space $K \times \Lambda$. It evolves by iteratively establishing that $\inf_K \sup_{z\in Z} D(z, \kappa, \lambda, \mathbb{P})$ is continuous in λ, which then implies that it has a maximizer. Specifically, $D(z, \kappa, \lambda, \mathbb{P})$ is continuous in (z, κ, λ) (w.r.t. the product of the Euclidean topology on $\mathbb{R}$, l_1 on K, Λ, respectively), due to the continuity of $(z - \sum_{i=0}^{\infty} \kappa_i X_{t,i})_+ - (z - \sum_{i=0}^{\infty} \lambda_i X_{t,i})_+$, Assumption 1 and Dominated Convergence. The CMT implies that $\sup_{z\in Z} D(z, \kappa, \lambda, \mathbb{P})$ is continuous in (κ, λ). We have that $K \underset{\text{SSD}}{\not\succeq} \Lambda$ iff $\exists \lambda^\star \in \Lambda - K$ such that $\forall \kappa \in K$, $\sup_{z\in Z} D(z, \kappa, \lambda^\star, \mathbb{P}) > 0$. The compactness of K and the continuity of $\sup_{z\in Z} D(z, \kappa, \lambda^\star, \mathbb{P})$ on the second argument imply that the latter holds iff $\inf_K \sup_{z\in Z} D(z, \kappa, \lambda^\star, \mathbb{P}) > 0$. The compactness of K also implies via Theorem 3.4 of Molchanov (2006) that $\inf_K \sup_{z\in Z} D(z, \kappa, \lambda, \mathbb{P})$ is continuous w.r.t. its third argument. Hence, $\inf_K \sup_{z\in Z} D(z, \kappa, \lambda^\star, \mathbb{P}) > 0$ is equivalent to $\sup_\Lambda \inf_K \sup_{z\in Z} D(z, \kappa, \lambda, \mathbb{P}) > 0$. □

Proof of Lemma 2. Follows by Lemma 1 and the monotonicity of Λ as a function of p. $\square$

Proof of Lemma 3. The proof evolves in the following steps: (i) we majorize $\sup_{z\in Z} D(z,\kappa,\lambda,\mathbb{P})$ by the supremum of $\int_Z D(z,\kappa,\lambda,\mathbb{P})\,d\cdot$ w.r.t. a set of linear operators. (ii) we validate a max-min result to interchange the order of optimization operators for $\inf_K \sup_{\cdot} \int_Z D(z,\kappa,\lambda,\mathbb{P})\,d\cdot$. (iii) we use an appropriate topology for $\mathcal{L}_{p,q}$ and establish appropriate continuity and generalized convexity properties for $\inf_K \sup_{F\in\mathcal{P}(Z)} \int_Z D(z,\kappa,\lambda,\mathbb{P})\,dF(z)$ as a function on $\Lambda\times\mathcal{L}_{p,q}$, so that we validate a max-min result to interchange the order of the outer pair of optimization operators in $\inf_{\mathcal{L}_{p,q}} \sup_{\Lambda} \sup_{\cdot} \inf_K \int_Z D(z,\kappa,\lambda,\mathbb{P})\,d\cdot$. (iv) Analogously to (iii), we validate a max-min result to interchange the order of the middle pair of optimization operators in $\sup_{\Lambda} \inf_{\mathcal{L}_{p,q}} \sup_{\cdot} \inf_K \int_Z D(z,\kappa,\lambda,\mathbb{P})\,d\cdot$. (v) We finally use the extreme point properties of the set of linear operators in (i) and the max-min inequality to obtain the result. Specifically, for (i) consider the space $\mathcal{P}(Z)$ comprised by the probability distributions that are supported on Z, and equipped with the weak topology. The space is convex and contains the degenerate distributions on the elements of Z as its extreme points. Then, by Theorem 15.9 of Aliprantis and Border (2006) we deduce that $\sup_{z\in Z} D(z,\kappa,\lambda,\mathbb{P}) \leq \sup_{F\in\mathcal{P}(Z)} \int_Z D(z,\kappa,\lambda,\mathbb{P})\,dF(z)$. For (ii), we have that due to Assumption 1, and the Lipschitz continuity property of $(\cdot)_+$, we have that $\sup_{Z,\Lambda^2} |D(z,\kappa,\lambda,\mathbb{P})| \leq 2\max_i \mathbb{E}[|X_i|] < +\infty$, hence the linear functional $F \to \int_Z D(z,\kappa,\lambda,\mathbb{P})\,dF(z)$ is also continuous w.r.t. F for all κ,λ, due to the Portmanteau Lemma. Furthermore, $\mathbb{E}\left[(z-\sum_{i=0}^{\infty}\kappa_i X_{t,i})_+\right]$ is convex in κ, due to the convexity and monotonicity of $(\cdot)_+$ and the linearity of $z - \sum_{i=0}^{\infty}\kappa_i X_{t,i}$ w.r.t. κ. Hence, since $K\in\mathcal{L}_{p,q}$ is closed and Λ is compact, the dual version of the Kneser-Fan Theorem (see Theorem 4.2' of Sion (1958)) implies that

$$\inf_K \sup_{F\in\mathcal{P}(Z)} \int_Z D(z,\kappa,\lambda,\mathbb{P})\,dF(z) = \sup_{F\in\mathcal{P}(Z)} \inf_K \int_Z D(z,\kappa,\lambda,\mathbb{P})\,dF(z).$$

For (iii), equip $\mathcal{L}_{p,q}$ with the PK-topology (see Definition 3.1.4 of Klein and Thompson (1984)). Due to Theorem 4.3.4-5 of Klein and Thompson (1984) $\mathcal{L}_{p,q}$ is compact. Due to Theorem 3.4 of Klein and Thompson (1984) and the boundedness and continuity of $D(\cdot,\cdot,\cdot,\mathbb{P})$, the mapping $\inf_K \int_Z D(z,\kappa,\lambda,\mathbb{P})\,dF(z) : \mathcal{L}_{p,q}\times\Lambda \to \mathbb{R}$ is jointly continuous for all F. Then, the boundedness of $D(\cdot,\cdot,\cdot,\mathbb{P})$ and the CMT imply that $\sup_{F\in\mathcal{P}(Z)} \inf_K \int_Z D(z,\kappa,\lambda,\mathbb{P})\,dF(z) :$ $\mathcal{L}_{p,q}\times\Lambda \to \mathbb{R}$ is also jointly continuous.

For any $t\in(0,1)$ and any $K_1, K_2 \in \mathcal{L}_{p,q}$, we have that

$$\begin{array}{c} t\inf_{K_1}\int_Z D(z,\kappa,\lambda,\mathbb{P})\,dF(z) + (1-t)\inf_{K_2}\int_Z D(z,\kappa,\lambda,\mathbb{P})\,dF(z) \\ \geq \min\left[\inf_{K_i^\star}\int_Z D(z,\kappa,\lambda,\mathbb{P})\,dF(z), i=1,2\right] \end{array},$$

where $K_i^\star$ is any element of $\mathcal{L}_{p,q}$ of support q that contains K_i, $i=1,2$. Analogously, we obtain from the previous and the monotonicity of sup

$$\begin{array}{c} t\sup_{F\in\mathcal{P}(Z)}\inf_{K_1}\int_Z D(z,\kappa,\lambda,\mathbb{P})\,dF(z) + (1-t)\sup_{F\in\mathcal{P}(Z)}\inf_{K_2}\int_Z D(z,\kappa,\lambda,\mathbb{P})\,dF(z) \\ \geq \min_{i=1,2}\inf_{K_i^\star}\int_Z D(z,\kappa,\lambda,\mathbb{P})\,dF(z) \\ \sup_{F\in\mathcal{P}(Z)}\left[t\inf_{K_1}\int_Z D(z,\kappa,\lambda,\mathbb{P})\,dF(z) + (1-t)\inf_{K_2}\int_Z D(z,\kappa,\lambda,\mathbb{P})\,dF(z)\right] \geq \\ \sup_{F\in\mathcal{P}(Z)}\min_{i=1,2}\inf_{K_i^\star}\int_Z D(z,\kappa,\lambda,\mathbb{P})\,dF(z) \end{array},$$

and the previous pair of displays implies that the mapping $\inf_K \int_Z D(z,\kappa,\lambda,\mathbb{P})\,dF(z) :$ $\mathcal{L}_{p,q} \to \mathbb{R}$ is convex-like for all (F,λ), and the mapping $\sup_{F\in\mathcal{P}(Z)}\inf_K \int_Z D(z,\kappa,\lambda,\mathbb{P})\,dF(z) :$ $\mathcal{L}_{p,q} \to \mathbb{R}$ is convex-like for all λ (see Section 2 of Sion (1958)). For any $t\in(0,1)$ and any $\lambda_1,\lambda_2\in\Lambda$ we have that due to Theorem 15.9 of Aliprantis and Border (2006)

$$\begin{array}{c} t\sup_{F\in\mathcal{P}(Z)}\inf_K\int_Z D(z,\kappa,\lambda_1,\mathbb{P})\,dF(z) + (1-t)\sup_{F\in\mathcal{P}(Z)}\inf_K\int_Z D(z,\kappa,\lambda_2,\mathbb{P})\,dF(z) \\ = t\sup_{z\in Z}\inf_K D(z,\kappa,\lambda_1,\mathbb{P}) + (1-t)\sup_{z\in Z}\inf_K D(z,\kappa,\lambda_2,\mathbb{P}) \end{array},$$

and the rhs of the previous display is less than or equal to $\max_\lambda \sup_{z\in Z}\inf_K D(z,\kappa,\lambda,\mathbb{P})$ and the maximum exists due to the joint continuity and boundedness of $D(\cdot,\cdot,\cdot,\mathbb{P})$, the CMT and the compactness of Λ. Hence, the mapping $\sup_{F\in\mathcal{P}(Z)}\inf_K\int_Z D(z,\kappa,\lambda,\mathbb{P})\,dF(z) : \Lambda \to \mathbb{R}$ again see Section 2 of Sion (1958)).

For (iv), for any $t \in (0,1)$ and any $F_1, F_2 \in \mathcal{P}(Z)$, we have that

$$\begin{aligned} t \inf_K \int_Z D(z,\kappa,\lambda,\mathbb{P})\, dF_1(z) + (1-t) \inf_K \int_Z D(z,\kappa,\lambda,\mathbb{P})\, dF_2(z) \\ \geq \inf_K \int_Z D(z,\kappa,\lambda,\mathbb{P})\, d\left[tF_1(z) + (1-t) F_2(z)\right] \end{aligned},$$

and thereby the mapping $\inf_K \int_Z D(z,\kappa,\lambda,\mathbb{P})\, dF(z) : \mathcal{P}(Z) \to \mathbb{R}$ is concave and hence concave-like for all $K \in \mathcal{L}_{p.q}$ and λ. Using the previous and applying twice the dual version of the Kneser-Fan Theorem, we jointly obtain the required results in steps (iii)-(iv) as,

$$\begin{aligned} \inf_{\mathcal{L}_{p,q}} \sup_\Lambda \sup_{F\in\mathcal{P}(Z)} \inf_K \int_Z D(z,\kappa,\lambda,\mathbb{P})\, dF(z) &= \sup_\Lambda \inf_{\mathcal{L}_{p,q}} \sup_{F\in\mathcal{P}(Z)} \inf_K \int_Z D(z,\kappa,\lambda,\mathbb{P})\, dF(z) \\ &= \sup_\Lambda \sup_{F\in\mathcal{P}(Z)} \inf_{\mathcal{L}_{p,q}} \inf_K \int_Z D(z,\kappa,\lambda,\mathbb{P})\, dF(z). \end{aligned}$$

Finally, for (v), again due to Theorem 15.9 of Aliprantis and Border (2006)

$$\sup_\Lambda \sup_{F\in\mathcal{P}(Z)} \inf_{\mathcal{L}_{p,q}} \inf_K \int_Z D(z,\kappa,\lambda,\mathbb{P})\, dF(z) = \sup_\Lambda \sup_{z\in Z} \inf_{\mathcal{L}_{p,q}} \inf_K D(z,\kappa,\lambda,\mathbb{P}).$$

The result follows by the max-min inequality. □

Proof of Proposition 1. K does not solve $\sup_{z\in Z} \sup_\Lambda \inf_{\mathcal{L}_{p,q}} \inf_K D(z,\kappa,\lambda,\mathbb{P})$ iff $\sup_{z\in Z} \sup_\Lambda \inf_K D(z,\kappa,\lambda,\mathbb{P}) > M(\Lambda, \mathcal{K}_{p,q}, \mathbb{P})$. Using the same argument as in the proof of Lemma 3 the latter is equivalent to $\sup_{F\in\mathcal{P}(Z)} \sup_\Lambda \inf_K \int_Z D(z,\kappa,\lambda,\mathbb{P})\, dF(z) > M(\Lambda, \mathcal{L}_{p,q}, \mathbb{P})$. Now, due to Fubini's Theorem we have that

$$\begin{aligned} &\sup_{F\in\mathcal{P}(Z)} \sup_\Lambda \inf_K \int_Z D(z,\kappa,\lambda,\mathbb{P})\, dF(z) \\ &= \sup_{F\in\mathcal{P}(Z)} \sup_\Lambda \inf_K \int_Z \mathbb{E}\left[\left(z - \textstyle\sum_{i=0}^\infty \kappa_i X_i\right)_+\right] - \mathbb{E}\left[\left(z - \textstyle\sum_{i=0}^\infty \lambda_i X_i\right)_+\right] dF(z) \\ &= \sup_{F\in\mathcal{P}(Z)} \sup_\Lambda \inf_K \int_Z \mathbb{E}\left(\left[\min\left(0, \textstyle\sum_{i=0}^\infty \lambda_i X_i - z\right) - \min\left(0, \textstyle\sum_{i=0}^\infty \kappa_i X_i - z\right)\right]\right) dF(z) \\ &= \sup_{F\in\mathcal{P}(Z)} \sup_\Lambda \inf_K \mathbb{E}\left[\int_Z \min\left(0, \textstyle\sum_{i=0}^\infty \lambda_i X_i - z\right) - \min\left(0, \textstyle\sum_{i=0}^\infty \kappa_i X_i - z\right) dF(z)\right] \\ &= \sup_{F\in\mathcal{P}(Z)} \left[\sup_\Lambda \mathbb{E}\left(u_F\left(\textstyle\sum_{i=0}^\infty \lambda_i X_i\right)\right) - \sup_K \mathbb{E}\left(u_F\left(\textstyle\sum_{i=0}^\infty \kappa_i X_i\right)\right)\right], \end{aligned}$$

and the result follows. □

Proof of Theorem 1. (a) we use the Ergodic Theorem uniformly in λ and continuously in z. Specifically, we derive the limiting behavior of $\frac{1}{T}\sum_{t=0}^{T}\left(z-\sum_{i=0}^{\infty}\lambda_i X_{t,i}\right)_+$ from the locally uniform in z and uniform in λ, version of the Ergodic Theorem applied on $\frac{1}{T}\sum_{t=0}^{T}\left(z-\sum_{i=0}^{\infty}\lambda_i X_{t,i}\right)_+$, noting that it is applicable due to Assumption 1 and the l_1 boundedness of Λ_∞, which imply that the rhs bound in (**??**) is independent of λ. Continuously uniform convergence then implies continuous hypo-convergence by Molchanov (2006).

For (b)-(c), (i) we establish that the associated set of functions has an integrable envelope, (ii) we use the fact that the associated sets of functions-which admit generalized derivatives w.r.t. the sample arguments-are bounded subsets of a weighted Sobolev space, and thus have controllable bracketing entropy numbers, and (iii) we use the above and the time series properties of X to verify the validity of an appropriate FCLT or maximal inequality. For (i) we have that due to Jensen's inequality,

$$\mathbb{E}\left[\sup_{z,\kappa}\left(\mu^T\begin{pmatrix}\left(z-\sum_{i=0}^{\infty}\kappa_i X_i\right)_+-\left(z-\sum_{i=0}^{\infty}\lambda_i X_i\right)_+\\ X^T\mathbb{I}\left\{z\geq\sum_{i=0}^{\infty}\lambda_i X_i\right\}(\kappa-\lambda)\end{pmatrix}\right)^{2+\varepsilon}\right]\leq$$
$$C\mathbb{E}\left[\left(\sup_{z,\kappa}\left(\left(z-\sum_{i=0}^{\infty}\kappa_i X_i\right)_+-\left(z-\sum_{i=0}^{\infty}\lambda_i X_i\right)_+\right)\right)^{2+\varepsilon}\right]+$$
$$C\mathbb{E}\left[\left(\sup_{z,\kappa}\left(X^T\mathbb{I}\left\{z\geq\sum_{i=0}^{\infty}\lambda_i X_i\right\}(\kappa-\lambda)\right)\right)^{2+\varepsilon}\right]\leq$$

$$\begin{gathered}C\mathbb{E}\left[\left(\sup_{z,\kappa}\left(\left(\sum_{i=0}^{\infty}(\lambda_i-\kappa_i)X_{t,i}\right)\right)\right)^{2+\varepsilon}\right]+\\ C\mathbb{E}\left[\max_i\left(X_i\mathbb{I}\left\{z\geq\sum_{i=0}^{\infty}\lambda_i X_i\right\}\right)^{2+\varepsilon}\right]\leq\\ 2^{1+\varepsilon}C\left(\mathbb{E}\left[\left(\sup_{z,\kappa}\left(\left(\sum_{i=0}^{\infty}\lambda_i X_i\right)\right)\right)^{2+\varepsilon}\right]+\mathbb{E}\left[\left(\sup_{z,\kappa}\left(\left(\sum_{i=0}^{\infty}\kappa_i X_i\right)\right)\right)^{2+\varepsilon}\right]\right)+\\ C\mathbb{E}\left[\max_i\left(|X_i|\right)^{2+\varepsilon}\right]\leq 2^{1+\varepsilon}C\mathbb{E}\left[\max_i\left(X_i\right)^{2+\varepsilon}\right]<+\infty.\end{gathered}\tag{22}$$

For (ii) , we have that for any $l\geq 1$, $\delta>0$, the function classes

$$\mathcal{M}_1:=\left\{\mathbb{R}^{\lfloor q(\ln T+1)\rfloor}\ni x\to\left(z-x^T\lambda\right)_+-\left(z-x^T\kappa\right)_+\right\}$$

and

$$\mathcal{M}_2 := \left\{ \mathbb{R}^{\lfloor q(\ln T+1)\rfloor} \ni x \to x^T \mathbb{I}\left\{ z \geq \sum_{i=0}^{\lfloor q(\ln T+1)\rfloor} \kappa_i^* x_i \right\} (\kappa - \kappa^*), z, \kappa, \kappa^* \right\}$$

are bounded subsets of the weighted Sobolev space $H_l^1\left(\mathbb{R}^{\lfloor q(\ln T+1)\rfloor}, \langle x\rangle^{2+\delta}\right)$ (this is the semi-normed space $\left\{ f : \mathbb{R}^{\lfloor q(\ln T+1)\rfloor} \to \mathbb{R}, \begin{array}{c} \|f\|_{l,2+\delta,\mu} := \\ \left(\int_{\mathbb{R}^{\lfloor q(\ln T+1)\rfloor}} \left[\left|\frac{f(x)}{(1+\|x\|)^{2+\delta}}\right|^l + \left|D\frac{f(x)}{(1+\|x\|)^{2+\delta}}\right|^l\right] d\mu\right)^{1/l} < +\infty \end{array} \right\}$, where D denotes partial derivation in the sense of distributions, and μ denotes the Lebesgue measure on $\mathbb{R}^{\lfloor q(\ln T+1)\rfloor}$-see 3.3.2 of Nickl and Potcher (2007)), due to the l_1-boundedness of K. In the notation of the aforementioned paper, choosing l such that $\frac{\lfloor q(\ln T+1)\rfloor}{l} \to 0$, $r = 2+\varepsilon$ and $\gamma = 3+\delta$, $\beta = 2+\delta$, $\mathfrak{M}$ the set of finite dimensional distributions of X^∞, we have that, due to Corollary 4.2 of Nickl and Potcher (2007), and for large enough T, the bracketing entropy of $\mathcal{M}$, as a function of $\epsilon > 0$, is universally bounded from above by $\ln c + \lfloor q(\ln T+1)\rfloor \ln\left(\frac{1}{\epsilon}\right)$ for some universal constant $c > 0$.

Then, from (i) above, (ii) the fact that $\beta_k \sim b^k$, and (iii) the fact that that the class have an $L^{2+\varepsilon}(\mathbb{P})$-integrable envelope due to (22), we get that Theorems 1 and 2 of Doukhan, Massart, and Rio (1995) are applicable and the results in (b) and (c) follow since $\frac{\ln\left(\binom{p}{\lfloor q(\ln T+1)\rfloor}\right)}{\sqrt{T}} + \frac{\lfloor q(\ln T+1)\rfloor}{\sqrt{T}} \to 0$. The latter holds since $\frac{\ln p}{\sqrt{T}} \to 0$ via Stirling's approximation on factorials and first order Taylor expansions on the logarithms. □

Proof of Theorem 2. The proof works by (i) establishing that the empirical LPM, $\frac{1}{T}\sum_{t=0}^{T}\left(z - \sum_{i=0}^{\infty}\kappa_i X_{t,i}\right)_+$, satisfies uniformly over z, w.h.p. the weak sub-modularity property of Elenberg et al. (2018), so that (ii) the guarantees results on the Forward Selection algorithm of the aforementioned paper hold w.h.p.. We do so by (iii) establishing that the first order Taylor expansion restricted on appropriate parts of the empirical LPM, is approximated by the analogous expansion of its population counterpart, uniformly over z, w.h.p.. Given (i), the statistical guarantees for the overall optimization problem follow (iv) by standard results on approximation of optimization problems and the CMT.

We first recall some notation mainly from convex analysis. Specifically, in what follows ∂

denotes the sub-gradient of an arbitrary real valued convex function defined on a locally convex space (see Ch. D of Hiriart-Urruty and Lemaréchal (2004)-HUL). Besides, for $\mathbb{Q} := \mathbb{P}, \mathbb{P}_T$ and $\mathbb{E}_{\mathbb{Q}}$ denoting integration w.r.t. $\mathbb{Q}$, $\mathbb{E}_{\mathbb{Q}}\left[\left(z - \sum_{i=0}^{\infty} \kappa_i X_{0,i}\right)_+\right]$ is convex in the second argument due to the convexity and monotonicity of $(\cdot)_+$ and the linearity of $z - \sum_{i=0}^{\infty} \kappa_i X_{t,i}$ w.r.t. κ.

Then, for any $\kappa \in \Lambda$ we have that $g_{z,T}(\kappa) := \frac{1}{T} \sum_{t=0}^{T} X_t \mathbb{I}_{z \geq \sum_{i=0}^{\infty} \kappa_i X_{0,i}+} \in \partial \mathbb{E}_{\mathbb{P}_T}\left(z - \sum_{i=0}^{\infty} \kappa_i X_{0,i}\right)_+$ due to Theorems 4.1.1. and 4.2.1 of HUL. Furthermore, due to Theorem 1 of Savare (1996) and the fact that X_t has a continuous density, we have that $g_z(\kappa) := \frac{\partial \mathbb{E}\left(z - \sum_{i=0}^{\infty} \kappa_i X_{0,i}\right)_+}{\partial \kappa} = \mathbb{E}\left[X_t \mathbb{I}_{z \geq \sum_{i=0}^{\infty} \kappa_i X_{t,i}}\right]$. Then, for any $(\kappa^\star, \kappa) \in \Lambda_{(\lfloor q(\ln T+1)\rfloor)}$, define the Taylor expansions $\mathcal{E}_{z,T}(\kappa^\star, \kappa) := \frac{1}{T} \sum_{t=0}^{T} \left(z - \sum_{i=0}^{\infty} \kappa_i^\star X_{t,i}\right)_+ - \frac{1}{T} \sum_{t=0}^{T} \left(z - \sum_{i=0}^{\infty} \kappa_i X_{t,i}\right)_+ - (\kappa^\star - \kappa)' g_{z,T}(\kappa)$, and $\mathcal{E}_z(\kappa^\star, \kappa) := \mathbb{E}\left[\left(z - \sum_{i=0}^{\infty} \kappa_i^\star X_{t,i}\right)_+\right] - \mathbb{E}\left[\left(z - \sum_{i=0}^{\infty} \kappa_i X_{t,i}\right)_+\right] - (\kappa^\star - \kappa)' g_z(\kappa)$.

Working towards (iii), and using the previous theorem, we have that for any $\delta > 0$ and any $C > 0, 0 < \epsilon < \frac{1}{4}$:

$$\begin{aligned}
&\mathbb{P}\left(\sup_z \sup_{\Lambda_{(\lfloor q(\ln(T+1))\rfloor)}, \|\kappa-\kappa^\star\| > \frac{C}{T^\epsilon}} \left(\frac{1}{\|\kappa-\kappa^\star\|^2} \left|\mathcal{E}_{z,T}(\kappa, \kappa^\star) - \mathcal{E}_z(\kappa, \kappa^\star)\right|\right) \geq \delta\right) \\
&\leq \mathbb{P}\left(\sup_z \sup_{\Lambda_{(\lfloor q(\ln(T+1))\rfloor)}, \|\kappa-\kappa^\star\| > \frac{C}{T^\epsilon}} \left(T^{2\epsilon} \left|\mathcal{E}_{z,T}(\kappa, \kappa^\star) - \mathcal{E}_z(\kappa, \kappa^\star)\right|\right) \geq \frac{\delta}{C}\right) \\
&\leq \mathbb{P}\left(\sup_z \sup_{\Lambda_{(\lfloor q(\ln(T+1))\rfloor)}} \left|\sqrt{T} D\left(z, \kappa, \kappa^\star, \mathbb{P}_T - \mathbb{P}\right)\right| \geq \frac{2\delta T^{\frac{1}{2}-2\epsilon}}{3C}\right) \\
&\quad + \mathbb{P}\left(\sup_z \sup_{\Lambda_{(\lfloor q(\ln(T+1))\rfloor)}} \left|G_T(z, \kappa, \kappa^\star)\right| \geq \frac{\delta T^{\frac{1}{2}-2\epsilon}}{3C}\right) = o(1),
\end{aligned} \tag{23}$$

where the final equality in the previous display follows from the first two parts of 1, the Lipschitz property of D w.r.t. the parameters, the fact that $T^{\frac{1}{2}-2\epsilon} \to +\infty$, and the Portmanteau Theorem.

Due to the bounds on the eigenvalues of $\mathbb{E}\left(z - \sum_{i=0}^{\infty} \kappa_i X_{0,i}\right)_+$ of Assumption 2, Theorem 6.1.2 of HUL, Paragraph 1.3.(d) in Ch. 4 of Hiriart-Urruty and Lemaréchal (2013), and the dual form of Remark 1 of Elenberg et al. (2018), we have that uniformly w.r.t. z and for any $(\kappa^\star, \kappa) \in \Lambda_{(\lfloor q(\ln T+1)\rfloor)}$, $\frac{m_{\lfloor q(\ln(T+1))\rfloor}}{2} \|\kappa - \kappa^\star\|^2 \leq \mathcal{E}_{z,T}(\kappa^\star, \kappa) \leq \frac{M_{\lfloor q(\ln(T+1))\rfloor}}{2} \|\kappa - \kappa^\star\|^2$. Due to this, and (23), uniformly w.r.t. z and for any $(\kappa^\star, \kappa) \in \Lambda_{(\lfloor q(\ln T+1)\rfloor)} \cap \left\{\|\kappa - \kappa^\star\| > \frac{C}{T^\epsilon}\right\}$,

$$\frac{m_{\lfloor q(\ln(T+1))\rfloor} + o_p(1)}{2} \|\kappa - \kappa^\star\|^2 \le \mathcal{E}_{z,T}(\kappa^\star, \kappa) \le \frac{M_{\lfloor q(\ln(T+1))\rfloor} + o_p(1)}{2} \|\kappa - \kappa^\star\|^2, \text{ w.h.p.,}$$

where the $o_p(1)$ terms are independent of z, λ. Thus (i) is established.

Then, for (ii), by noting that Theorem 1 of Elenberg et al. (2018) is also valid if the gradient in its proof is substituted by any fixed element of the sub-gradient, and using the previous display, the inclusion $\Lambda_{(\lfloor q(\ln T+1)\rfloor)} \cap \left\{\|\kappa - \kappa^\star\| > \frac{C}{T^\epsilon}\right\} \subseteq \Lambda_{(\lfloor q(\ln T+1)\rfloor)}$ and the discussion immediately after Remark 1 of Elenberg et al. (2018), we get that w.h.p. $\mathcal{K}^{\text{FS}}(\Lambda, \mathcal{L}_{p,q}, z, \mathbb{P}_T, q\ln T) \le \left(1 - \frac{1}{T^{\gamma_T}}\right) \inf_{\mathcal{L}_{p,q}} \inf_K \frac{1}{T}\sum_{t=0}^T \left(z - \sum_{i=0}^\infty \kappa_i X_{t,i}\right)_+$, where $\gamma_T := \frac{m_{\lfloor q(\ln(T+1))\rfloor} + o_p(1)}{M_{\lfloor q(\ln(T+1))\rfloor} + o_p(1)}$, establishing (ii). Finally, working towards (iv), note that Assumption 2, Theorem 1, the PK-convergence of $\Lambda_{(\lfloor q(\ln T+1)\rfloor)} \cap \left\{\|\kappa - \kappa^\star\| > \frac{C}{T^\epsilon}\right\}$ to $\Lambda_{(\lfloor q(\ln T+1)\rfloor)}$, and the CMT imply then (4). The final result follows from the dual version of Theorem 3.4 (Ch. 5, p. 338) of Molchanov (2006) and the CMT. □

Proof of Theorem (3). The strategy of the proof evolves as: (i) we establish the existence of a further subset of the above mentioned parameter set, which is compact and also contains the part of the population optimizers associated with non-degeneracy of the limiting empirical process, as well as the analogous empirical optimizers w.h.p.. (ii) we use the compactness of the aforementioned set to apply the generalized Delta method on the restricted empirical process.

For (i), first, due to Theorem 1 of Elenberg et al. (2018), the results of Theorem 2 are valid since $r = q(\ln T)^2$. Using additionally CM, the final result of Theorem 1 and Theorem 3.4 (Ch. 5, p. 338) of Molchanov (2006), we also have the approximation $\sqrt{T}\left|\inf \frac{1}{T}\sum_{t=0}^T \left(z - \sum_{i=0}^\infty \kappa_i X_{i,t}\right)_+ - \inf_{\text{csupp}(\kappa)\le q} \frac{1}{T}\sum_{t=0}^T \left(z - \sum_{i=0}^\infty \kappa_i X_{i,t}\right)_+\right| = o_p(1)$ where the remainder is independent of z and the first empirical infimum is derived via forward selection.

Then the limiting behavior of the empirical process $\sqrt{T}D(z, \kappa, \lambda, \mathbb{P}_T - \mathbb{P})$ restricted on

$Z \times \Lambda_{\infty} \times \tilde{\Lambda}_{(\lfloor q(\ln T+1)\rfloor)}$ is obtained by the (b) part of Theorem 1. Now, for (i), the proof of Lemma 3 and CO, imply that $\sup_{\Lambda_{\infty}} \inf_{\text{csupp}(\kappa)\leq q} \int_{Z} D(z,\kappa,\lambda,\mathbb{P})\, dF(z) : \mathcal{P}(Z) \to \mathbb{R}$ is strictly concave on $Z - \{\inf(Z)\}$. Hence, the set of optimizers $\arg\max_{Z-\{\inf(Z)\}} \sup_{\Lambda_{\infty}} \inf_{\text{csupp}(\kappa)\leq q} D(z,\kappa,\lambda,\mathbb{P})$ is singleton. Thereby, Theorem 3.4 of Molchanov (2006) implies that when $M(\Lambda_{\infty},\mathcal{L}_{\infty,q},\mathbb{P}) > 0$, Γ is compact , and when $M(\Lambda_{\infty},\mathcal{L}_{\infty,q},\mathbb{P}) = 0$, $\Gamma - \left(\{\inf Z\} \times \Lambda_{\infty} \times \tilde{\Lambda}_{(q)}\right)$ is compact. Hence and due to Assumption 2, there exists some $\varepsilon > 0$ for which Γ_{ε}, i.e., the set of triples from $Z \times \Lambda_{\infty} \times \tilde{\Lambda}_{(q)}$ of infimum distance from $\Gamma_{\star}$, less than or equal to ε, is non-empty compact. Here, $\Gamma_{\star}$ equals Γ when $M(\Lambda_{\infty},\mathcal{L}_{\infty,q},\mathbb{P}) > 0$, and equals $\Delta := \left(\Gamma - \left(\{\inf Z\} \times \Lambda_{\infty} \times \tilde{\Lambda}_{(q)}\right)\right) \cup \left(\{\inf Z\} \times \Lambda^{\star}_{\infty} \times \tilde{\Lambda}^{\star}_{(q)}\right)$ when $M(\Lambda_{\infty},\mathcal{L}_{\infty,q},\mathbb{P}) = 0$, where $\Lambda^{\star}_{\infty} \times \tilde{\Lambda}^{\star}_{(q)}$ is the compact set comprised by the (λ,κ) that appear in some triplet of Γ for $z > \inf Z$. The distance is the maximum between the Euclidean metric for the z parts of the triples, and the L_1 distances for the λ and κ parts. Due to Theorem 1 and Theorem 3.4 of Molchanov (2006), we have that Γ_{ε} contains solutions of $\sup_{Z\times\Lambda} \inf_{\text{csupp}(\kappa)\leq q} \frac{1}{T}\sum_{t=0}^{T}\left[\left(z - \sum_{i=0}^{\infty}\kappa_i X_{i,t}\right)_{+} - \left(z - \sum_{i=0}^{\infty}\lambda_i X_{i,t}\right)_{+}\right]$ w.h.p., and the solutions that it may miss correspond only to the case $M(\Lambda_{\infty},\mathcal{L}_{\infty,q},\mathbb{P}) = 0$ and at which the empirical optimization problem above is identically zero.

Finally, for (ii), we have that the result follows from Theorem 2, Theorem 3.4 of Molchanov (2006) and, Theorem 2.1 and Lemma B.1 of Fang and Santos (2014), restricting the optimizations appearing in the empirical process on Γ_{ε}, by noticing that when $M(\Lambda_{\infty},\mathcal{L}_{\infty,q},\mathbb{P}) = 0$, $\sup\inf_{\Delta} \mathcal{G}(z,\lambda,\kappa) = \sup\inf_{\Gamma} \mathcal{G}(z,\lambda,\kappa)$ due to the degeneracy at zero enforced by the elements of $\Gamma - \Delta$ on $\mathcal{G}$, and the fact that Δ already contains $\{\inf Z\} \times \Lambda^{\star}_{\infty} \times \tilde{\Lambda}^{\star}_{(q)}$ the elements of which also imply degeneracy at zero for the Gaussian process. □

Proof of Proposition 2. The proof proceeds as follows: (i) we establish the weak convergence of the scaled by b_T, discrepancy between the subsampling empirical process, evaluated at any convergent subsequence of the Forward Selection optimizers, and the population optimum, to the sup inf of the Gaussian process appearing in the previous result over $\Gamma^{\star}$. (ii) we establish conservativeness by showing that the cdf of the weak limit is continuous at its $1-\alpha$ quantile.

For (i), we have that from the weak convergence to the empirical process in the proof of Theorem 3, and applying Proposition 7.3.1 of Politis, Romano and Wolf (1999), we obtain that

$$\sqrt{b_T}\left(\mathbb{E}^{\star}\left[D\left(z,\kappa,\lambda,\mathbb{P}_{t,b_T}\right)\right]-D\left(z,\kappa,\lambda,\mathbb{P}_T\right)\right)\rightsquigarrow\mathcal{G}\left(z,\lambda,\kappa\right),$$

in $\ell^{\infty}\left(Z\times\Lambda_{\infty}\times\Lambda_{\infty}\right)$, where $\mathbb{E}^{\star}\left[\cdot\right]$ denotes expectation w.r.t. the empirical distribution of $D\left(z,\kappa,\lambda,\mathbb{P}_{t,b_T}\right)$ across $t=1,\ldots,T-b_T+1$.

In what follows, we also denote with (T) the index set of the subsequence of $\kappa_{z,T}$ associated with the examined accumulation point, for notational simplicity. Due to that (see the proof of Theorem 3), $\sqrt{T}\left|\inf\frac{1}{T}\sum_{t=0}^{T}\left(z-\sum_{i=0}^{\infty}\kappa_i X_{i,t}\right)_+-\inf_{\text{csupp}(\kappa)\leq q}\frac{1}{T}\sum_{t=0}^{T}\left(z-\sum_{i=0}^{\infty}\kappa_i X_{i,t}\right)_+\right|=o_p\left(1\right)$ uniformly in z, the definition of $\kappa_{z,T}$, and that $\frac{b_T}{T}\rightarrow0$, we have that $\sqrt{b_T}\left(\inf_{\text{csupp}(\kappa)\leq q}D\left(z,\kappa,\lambda,\mathbb{P}_T\right)-D\left(z,\kappa_{z,T},\lambda,\mathbb{P}_T\right)\right)=o_p\left(1\right)$ uniformly on $Z\times\Lambda_{\infty}$. It implies that $\sqrt{b_T}\left(\sup_{Z\times\Lambda}\inf_{\text{csupp}(\kappa)\leq q}D\left(z,\kappa,\lambda,\mathbb{P}_T\right)-\sup_{Z\times\Lambda}D\left(z,\kappa_{z,T},\lambda,\mathbb{P}_T\right)\right)=o_p\left(1\right)$. Employing a) the use of Skorokhod representations, applicable due to Theorem 3.7.25 of Giné and Nickl, (2016), b) the convergence above, c. Theorem 3.4 of Molchanov (2006), d) Theorem 2.1 and Lemma B.1 of Fang and Santos (2014) along with the compactness of Γ-working similarly to the proof of Theorem 3 with Γ_{ε}, e) the fact that $\left(\kappa_{z,T}\right)_z$ are optimizers of $\mathcal{K}^{\text{FS}}\left(\Lambda,\mathcal{L}_{p,q},z,\mathbb{P}_T,r_T\left(q\right)\right)-\mathcal{L}\left(\Lambda,z,\mathbb{P}_T\right)$, which due to Theorem 2 converges to the deterministic $\mathcal{K}\left(\Lambda^{\infty},\mathcal{L}_{\infty,q},z,\mathbb{P}\right)-\mathcal{L}\left(\Lambda_{\infty},z,\mathbb{P}\right)$, and thereby $\left(\kappa_{z,T}\right)_z$ are asymptotically independent to $\sqrt{b_T}\left(\sup_{Z\times\Lambda}\mathbb{E}^{\star}\left[D\left(z,\kappa_{z,T},\lambda,\mathbb{P}_{t,b_T}\right)\right]-M^{\text{FS}}\left(\Lambda,\mathcal{L}_{p,q},\mathbb{P}_T,q\left(\ln T\right)^2\right)\right)$, and f) the fact that $\frac{b_T}{T}\rightarrow0$, we obtain that

$$\sqrt{b_T}\left(\sup_{Z\times\Lambda}\mathbb{E}^{\star}\left[D\left(z,\kappa_{z,T},\lambda,\mathbb{P}_{t,b_T}\right)\right]-M^{\text{FS}}\left(\Lambda,\mathcal{L}_{p,q},\mathbb{P}_T,q\left(\ln T\right)^2\right)\right)\rightsquigarrow\sup\inf_{\Gamma^{\star}}\mathcal{G}\left(z,\lambda,\kappa\right).$$

For (ii), first, the definition of $\Gamma^{\star}$ implies that $\sup\inf_{\Gamma^{\star}}\mathcal{G}\left(z,\lambda,\kappa\right)\geq\sup\inf_{\Gamma}\mathcal{G}\left(z,\lambda,\kappa\right)$. Then conservativeness follows from this inequality as long as the cdf of $\sup\inf_{\Gamma^{\star}}\mathcal{G}\left(z,\lambda,\kappa\right)$ is continuous at its $1-\alpha$ quantile. From Lemma 18.15 of van der Vaart (2000), we have that for

$\mu, v \in \Gamma^\star$ and $\mathcal{G}_\mu, \mathcal{G}_v$ the Gaussian process $\mathcal{G}$ evaluated there,

$$0 \leq \sigma^2 := \sup_{\Gamma^\star} \mathbb{E}\left[\mathcal{G}_\mu^2\right] \leq \sup_{\mu,v\in\Gamma^\star} \mathbb{E}\left[(\mathcal{G}_\mu - \mathcal{G}_v]^2\right) < +\infty.$$

Hence due to the zero mean function of $\mathcal{G}_\mu$, and Furnique's inequality (see Relation (1,1) in Samorodnitsky (1991)), we have that for $0 < \varepsilon < 1$, there exists a $\kappa(\varepsilon)$, such that

$$\mathbb{E}\left[\sup_{\Gamma^\star} \mathcal{G}_\mu^2\right] = \int_0^{+\infty} \mathbb{P}\left(\sup_{\Gamma^\star} |\mathcal{G}_\mu| > \sqrt{y}\right) dy \leq 2\kappa(\varepsilon) \int_0^{+\infty} \exp\left(\frac{-(1-\varepsilon)}{2\sigma^2} y\right) dy < +\infty.$$

Then, Ch. 2 of Nualart (2006), (see the remark after the proof of Proposition 2.1.11 (p. 109)) implies the existence of the square integrable Malliavin derivative for $\mathcal{G}_\mu$. Nualart (2006) implies then that the Malliavin derivative of $\mathcal{G}_\mu$ equals zero only at trivial triples. The previous imply the validity of Assumption 1 of Arvanitis, Scaillet and Topaloglou (2019) for $\mathcal{T} = \{0\}$ in their notation, when trivial triples exist, and $\mathcal{T} = \emptyset$ when trivial triples do not exist. In the latter case Theorem 1 of Arvanitis, Scaillet and Topaloglou (2019), implies (7) for any $\alpha \in (0, 1)$. In the former case, ND assumes the existence of the non trivial $(\lambda^\star, \kappa^\star, z^\star) \in \Gamma^\star$ for which we have that,

$$\mathbb{P}\left(\sup\inf_{\Gamma^\star} \mathcal{G}(z, \lambda, \kappa) > 0\right) \geq \mathbb{P}\left(\sup\inf_{\Gamma^\star} \mathcal{G}(z, \lambda, \kappa^\star) > 0\right) \geq \mathbb{P}\left(\mathcal{G}(z^\star, \lambda^\star, \kappa^\star) > 0\right) = \frac{1}{2},$$

due to non-degeneracy and zero mean Gaussianity. The result then follows again from Theorem 1 of Arvanitis, Scaillet and Topaloglou (2019), and (8) follows from the previous by noting that in this special case, $\Gamma = \Gamma^\star$ due to Theorem 3.4 of Molchanov (2006). □