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

April 2026

Stelios Arvanitis

Olivier Scaillet

Nikolas Topaloglou

Venice Time Series Workshop
Ca' Foscari University of Venice
In the *La Serenissima*

The paper: Theoretical contributions

- We develop and implement methods for determining whether relaxing sparsity constraints on portfolios improves the investment opportunity set for risk-averse investors.
- We extend the notion of SSD-Spanning of Arvanitis et al. (2018) by allowing endogenous spanning sets in large dimensional settings.
- We formulate a new estimation procedure for sparse second-order stochastic spanning based on a greedy algorithm and Linear Programming.
- We derive the asymptotic greedy approximation ratio whether spanning holds or not.

The paper: Empirical findings

- From large equity datasets, we estimate the expected utility loss due to possible under-diversification, and find evidence that there is no benefit from expanding a sparse opportunity set beyond 30-40 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 10-15 assets in crisis periods.

The questions that we address

- Is it possible to build a sparse portfolio of *an exogenously given* "dimension" q from a large universe of assets of dimension p so that we cannot get further improvement from considering additional assets for any risk averter?
- This brings the analysis into the realm of Second Order Stochastic Dominance (SSD).
- If not, how much do we loose by limiting ourselves to this sparse portfolio in terms of the expected utilities?
- Can we design an optimization algorithm to compute this sparse portfolio from available data?
- Can we obtain some sort of asymptotic statistical guarantees?

Recap on SSD

- Probability measures are essentially methods of integrating measurable functions:
- $\mathbb{P},\ \mathbb{Q} \in \mathcal{P}(\mathbb{R})$, the set of Borel probability measures on the real line.
- $\mathcal{U}_2 := \{u : \mathbb{R} \to \mathbb{R}, u \text{ increasing and concave}\}$ represents the set of non-satiable and risk averse preferences inside the expected utility paradigm.
- $\mathbb{P} \underset{\text{SSD}}{\succeq} \mathbb{Q}$ iff $\mathbb{E}_{\mathbb{P}}(u) \geq \mathbb{E}_{\mathbb{Q}}(u),\ \forall u \in \mathcal{U}_2$.
- The pre-order is representable by the Russell-Seo sub-class of utilities: $\mathbb{P} \underset{\text{SSD}}{\succeq} \mathbb{Q}$ iff $\int (z-x)_+ (d\mathbb{P} - d\mathbb{Q}) \leq 0,\ \forall z \in \text{supp}(\mathbb{P}) \cup \text{supp}(\mathbb{Q})$.
- Integration by parts then implies a representation based on integrals of the respective cdfs, quantile functions, etc.

Probabilistic Framework

- Stationary p-dimensional financial returns process (X_t); $\mathbb{P}$ is the joint distribution of X_0. $\mathbb{P}_T$ is the empirical distribution of the available sample.
- The portfolio weights space is Λ the $p-1$ dimensional standard simplex.
- λ, κ are generic elements of Λ.
- The analysis allows that $p \to \infty$ with T; Λ converges to the $\mathbb{N}$- simplex.

Assumption (MS-Moments)

$$\text{For some } \varepsilon > 0, \max_i \mathbb{E}\left[|X_i|^{2+\varepsilon}\right] < +\infty.$$

- MS implies $D(z, \kappa, \lambda, \mathbb{P}) := \mathbb{E}\left[\left(z - \sum_{i=0}^{p} \kappa_i X_i\right)_+\right] - \mathbb{E}\left[\left(z - \sum_{i=0}^{p} \lambda_i X_i\right)_+\right]$ is bounded and continuous in z, λ, κ.

SSD Spanning I

Definition (SSD)

$$\kappa \underset{\text{SSD}}{\succeq} \lambda, \text{ iff } D(z, \kappa, \lambda, \mathbb{P}) \leq 0 \text{ for all } z.$$

Definition (SSD Spanning-Arvanitis et al. (2018))

$$(\Lambda \supset) K \text{ SSD spans } \Lambda, K \underset{\text{SSD}}{\succeq} \Lambda, \text{ iff } \forall \lambda \in \Lambda,\ \exists \kappa \in K : \kappa \underset{\text{SSD}}{\succeq} \lambda, \text{ i.e.}$$
$$\forall \lambda \in \Lambda,\ \exists \kappa \in K : \forall z, D(z, \kappa, \lambda, \mathbb{P}) \leq 0.$$

- $K \underset{\text{SSD}}{\succeq} \Lambda$ is equivalent to $M(\Lambda, K, \mathbb{P}) := \sup_{\lambda \in \Lambda} \inf_{\kappa \in K} \sup_z D(z, \kappa, \lambda, \mathbb{P}) \leq 0$,
- and invoking Sion's MinMax it is equivalent to $M(\Lambda, K, \mathbb{P}) = \sup_u (\sup_{\lambda \in \Lambda} \mathbb{E}(u(\lambda' X)) - \sup_{\kappa \in K} \mathbb{E}(u(\kappa' X))) = 0$.
- Spanning implies no diversification loss from restricting portfolio construction to K across all non-satiable risk averters.
- Thus $M(\Lambda, K, \mathbb{P})$ represents the maximal expected utility diversification loss across all risk averters from constraining the investment opportunities set to K.

SSD Spanning II

- It is provable that $K \underset{\text{SSD}}{\succeq} \Lambda$ iff K contains every un-dominated element of Λ (Pareto set);
- No spanning is thus equivalent to the existence of a (worst case) Pareto element in Λ with nothing equivalent in K.
- Any κ solution to the optimization problem defining $M(\Lambda, K, \mathbb{P})$ would approximate from inside K this in terms of expected utility.
- In Arvanitis et al. (2018) the spanning candidate K is exogenous.
- In the present high dimensional setting we ask something a bit more complicated: given a sparsity parameter $q \ll p$, does there exist some sub-simplex $K \subset \Lambda$, such that

  a any $\kappa \in K$ has at most q non-zero components, and,

  b $K \underset{\text{SSD}}{\succeq} \Lambda$.

q-Sparse SSD Spanning I

Definition (q-Sparse Spanning SSD)

Given q, there exists a sub-simplex $K \subset \Lambda$ with $\text{csupp}(K) \leq q$ and such that $K \underset{\text{SSD}}{\succeq} \Lambda$, i.e. $\forall \lambda \in \Lambda,\ \exists \kappa \in K : \kappa \underset{\text{SSD}}{\succeq} \lambda$.

- $\text{csupp}(K) := \max_{\kappa \in K} \#\{i : \kappa_i \neq 0\}$ is the portfolio set support.

Lemma (Functional Characterization)

$K \underset{\text{SSD}}{\succeq} \Lambda$ iff $M(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}) := \inf_{K \in \mathcal{L}_{p,q}} \sup_{\Lambda} \inf_{\kappa \in K} \sup_{z \in Z} D(z, \kappa, \lambda, \mathbb{P}) \leq 0$.

- $\mathcal{L}_{p,q} := \{K \subset \Lambda : K \text{ closed}, 0 < \text{csupp}(K) \leq q\}$;
  - sidenote: if $p > 2q$, $\mathcal{L}_{p,q}$ is realizable as a sub-simplex of the $p-1$ simplex.
- Searching over the q-supported sub-simplices of Λ induces another optimization layer $inf_{K \in \mathcal{L}_{p,q}}$.

q-Sparse SSD Spanning II

Lemma (Numerically Useful Functional Characterization)

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

- Proof by successive apps of Sion's MinMax, and use of a compact PK topology for $\mathcal{L}_{p,q}$.
- Allows for separation of the optimizations w.r.t. Λ and $\mathcal{L}_{p,q}\times K$, for any z.
- Useful when the outer optimization over $z\in Z\subseteq[c,+\infty)$ is approximated by discretization.

Lemma (Non-Solution)

$K\in\mathcal{L}_{p,q}$ does not solve $\sup_{\Lambda}\sup_{z\in Z}\inf_{\mathcal{L}_{p,q}}\inf_{K} D\left(z,\kappa,\lambda,\mathbb{P}\right)$, iff there exists some $\lambda\in\Lambda$ and some $u\in\mathcal{U}_2$ such that $D(u,\lambda,\kappa,\mathbb{P})>M\left(\Lambda,\mathcal{L}_{p,q},\mathbb{P}\right)$ for any $\kappa\in K$.

- For given q, sparse spanning does not hold iff there exists some SSD-Pareto element in Λ of support greater than q.

q-Sparse SSD Spanning III

- In this case:
  - $M\left(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}\right)$ represents the optimal under-diversification if portfolio choice is restricted to q-sparse portfolios,
  - the solution K achieves the optimality bound; it thus contains the q-sparse elements that best approximate the Pareto elements of greater support (if any).
- We can thus ask: how does $M\left(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}\right)$ vary with q? M is monotone, low global minima would imply sparse diversification.

Statistical and Computational Aspects

- But $M\left(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}\right)$ is latent.
- We estimate it via $M\left(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}_T\right)$.
- Feasible only if the associated empirical optimization problems are solvable.
- We essentially need a technology for solving the $\inf_{\mathcal{L}_{p,q}}$ problem.
  - We have to solve

    $$\mathcal{K}\left(\mathcal{L}_{p,q}, z, \mathbb{P}_T\right) := -\sup_{K \in \mathcal{L}_{p,q}} f(K),$$

    for the real set function

    $$\mathcal{L}_{p,q} \ni K \to f(K) := \sup_K -\frac{1}{T}\sum_{t=0}^{T}\left(z - \sum_{i=0}^{p}\kappa_i X_{i,t}\right)_+.$$

  - We use a greedy algorithm that performs an iterative features (i.e. "basis assets") extraction.

Computational Aspects: Greedy Algorithm+LP I

Algorithm (Forward Stepwise Selection-Elenberg et al. (2018))

Inputs: the sparsity Parameter $q < p$, the # of iterations $r(q)$, and the set function f:

a Choose the initial set $S_0 = \emptyset$,

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

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

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

- $[p]$ denotes the set of basis elements of the $p-1$ simplex.
- S_i is the subset of the components of X at iteration i; $K(S_i)$ is the sub-simplex of Λ formed by S_i.
- If $r(q) = q$, and $i = r(q)$, the algorithm returns $\mathcal{K}^{\text{FS}}\left(\Lambda, \mathcal{L}_{p,q}, z, \mathbb{P}_T, r(q)\right)$ that numerically approximates $\mathcal{K}\left(\Lambda, \mathcal{L}_{p,q}, z, \mathbb{P}_T\right)$, along with "features extraction".
- At each iteration, the algorithm adds one base asset (feature)—the one that yields the largest marginal reduction in downside risk (SSD loss)—thereby sequentially constructing the sparse portfolio.

Computational Aspects: Greedy Algorithm+LP II

- If f is (weakly) submodular (i.e. satisfies approximate diminishing returns), FS returns a solution that is provably within a constant factor of the optimum (Nemhauser, Wolsey and Fisher (1978)), and it turns out to be the best approximation ratio possible for the problem (Nemhauser and Wolsey (1978)).
- Elenberg et al. (2018) connect weak submodularity to conditions of Restricted Strong Convexity and Restricted Smoothness for the underlying criteria, with the submodularity ratio governed by curvature.
- RSC and RS are conditions that bound away from zero and infinity the eigenvalues of the Hessian in neighborhoods of the optimizers of sufficiently low metric entropy; imply $\mathcal{K}^{\text{FS}} \leq (1 + \exp(-\frac{m}{M}\frac{r(q)}{q}))\mathcal{K}$, with $m,\ M$ appropriate RSC and RS indices, while in our setting the simplicial structure of Λ effectively controls the RS component.
- The $\inf_\Lambda$ and $\inf_K$ parts are handled via standard LP approximations.
- The $\inf_Z$ part is approximated via discretization (could be enhanced with piecewise linear approximations of the Russell-Seo mixtures).
- FS+LP produce $M^{\text{FS}}\left(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}_T, r(q)\right)$ that numerically approximates $M\left(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}_T\right)$.

Computational Aspects: Greedy Algorithm+LP III

- To get an idea of the associated computational complexity in our empirical application:
  - **Utility Approximation:** 715 distinct concave piecewise-linear utility functions (from Russel-Seo ramps)
  - **LP Calls:**
    - 2 LP problems per utility function $\Rightarrow$ 715 LPs+ 715 LPs per iteration
    - FS adds one asset per iteration, for up to $q = 45$ assets
    - Total LP evaluations: $\sim 715 \times 46 =$ **32,890**
  - **Hardware Used:**
    - Desktop PC with 3.6 GHz Intel i7 (24-core), 128 GB RAM
    - MATLAB + GAMS with Gurobi solver
  - **Estimated Run-Time:** For full sparse SSD experiment: $\sim$ **50–80 minutes**
    - Including FS iterations, LP solving, and utility discretization

Empirical Sparse Optimization: Some Limit Theory I

- Interested in the asymptotic behavior of $M_\delta^{\mathrm{FS}}\left(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}_T, r(q)\right)$ as $p, T \to \infty$.
- Assume $z \in [c, c^\star]$ for simplicity.
- Also $\delta > 0$, leaves out the Russell-Seo functions that correspond to thresholds $c \leq z < c + \delta$; truncated SD compatible with discretization.
- $\Lambda_{(k)}$ denotes the subset of Λ comprised of vectors with supports bounded by k.

Assumption (Restricted Strong Convexity)

$0 < m_{(2q)}$ independent of p, with $m_{(2q)}$ denoting the infimum over $\Lambda_{(2q)} \times [c + \delta, +\infty)$, of the smallest eigenvalue of the Hessian matrix of $\mathbb{E}\left(z - \sum_{i=0}^{p} \kappa_i X_0^{(i)}\right)_+$ as a function of κ.

Assumption (Mixing)

(X_t) is strictly stationary and absolutely regular with mixing coefficients $(\beta_m)_{m \in \mathbb{N}}$ that satisfy $\beta_m \sim b^m$ for some $b \in (0, 1)$, independent of p, as $m \to \infty$.

Empirical Sparse Optimization: Some Limit Theory II

Theorem ($r(q) = q$)

Under the previous assumptions, if $\frac{\ln p}{\sqrt{T}} \to 0$, then,

$$M_\delta^{\mathrm{FS}}\left(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}_T, q\right) \rightsquigarrow M_\delta^\star,$$

with

$$M_\delta^\star \leq M_\delta\left(\Lambda_\infty, \mathcal{L}_{\infty,q}, \mathbb{P}\right) + \exp(-m_{(2q)}) \sup_{z \in [c+\delta, c^\star]} \mathcal{K}\left(\mathcal{L}_{\infty,q}, z, \mathbb{P}\right).$$

Furthermore:

$$\sqrt{T}\left(M_\delta^{\mathrm{FS}}\left(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}_T, q\right) - M_\delta^\star\right) \rightsquigarrow \sup \inf_{(z,\lambda,\kappa)\in\Gamma^\star} \mathcal{G}\left(z, \lambda, \kappa\right);$$

$\mathcal{G}\left(z, \lambda, \kappa\right)$ is a zero mean Gaussian process, and,
$\Gamma^\star := \arg\max_{z\in[c+\delta,c^\star],\lambda\in\Lambda_\infty}(\delta^\star - \arg\min_{\{\kappa:\mathrm{csupp}(\kappa)\leq q\}})D\left(z, \lambda, \kappa, \mathbb{P}\right)$, where
$\delta^\star - \arg\min$ denotes $\delta^\star$ approximate minimizers, for
$\delta^\star := \exp(-m_{(2q)})\mathcal{K}\left(\mathcal{L}_{\infty,q}, z, \mathbb{P}\right)$.

Empirical Sparse Optimization: Some Limit Theory III

- Proof Skeleton:
  - handle metric entropy arguments working on appropriate Sobolev spaces;
  - establish locally uniform LLNs and FCLT's for the sparse optimization so that extensions w.r.t. sub-differentials of the guarantees results of Elenberg et al. (2018) are applicable to the empirical functionals;
  - handle metric entropy arguments for FCLTs for the high dimensional optimization and then work via an extension of the generalized Delta method of Cárcamo et al. (2020).
- The greedy approximation ratios asymptotically persist and produce inconsistency;
- the limiting optimum is an approximate one due to the $\exp(-m_{(2q)})\sup_{z\in[c+\delta,c^\star]}\mathcal{K}\left(\mathcal{L}_{\infty,q},z,\mathbb{P}\right)$ optimization error.
- the second result can be used to construct asymptotically conservative "fast" subsampling CIs for $M_\delta^\star$.

Empirical Sparse Optimization: Some Limit Theory IV

- But what happens if we let $r(q) \to \infty$ with T?

Assumption (Restricted Strong Convexity*)

For some $\epsilon > 1$, $m_{(\lfloor q(\ln^{\epsilon}(T+1))\rfloor)} \ln^{\epsilon-1} T \to +\infty$ locally uniformly in $(c, c^{\star}]$.

- Allows for slowly varying asymptotic singularity.
- Sufficient condition: $\lambda_{\min}(V) \ln T \to 0$; V second moment matrix; proof via generalized derivatives.

Assumption (q-Population Sparcity)

Every portfolio optimizer that appears in the population criterion is q-sparse.

- Does not allow for equivalencies with "higher-dimensional" portfolios.
- With some more effort, it is then provable that:

Empirical Sparse Optimization: Some Limit Theory V

Theorem ($r(q) \to \infty$)

Under the modified assumptions, if $\frac{\ln p}{\sqrt{T}} \to 0$, then for fixed q,

$$M^{\text{FS}}\left(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}_T, q \ln^{\epsilon} T\right) \rightsquigarrow M\left(\Lambda_\infty, \mathcal{L}_{\infty,q}, \mathbb{P}\right),$$

and

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

$\mathcal{G}\left(z, \lambda, \kappa\right)$ is zero mean Gaussian, and,
$\Gamma := \arg\max_{z\in Z, \lambda\in\Lambda_\infty} \min_{\{\kappa:\text{csupp}(\kappa)\leq q\}} D\left(z, \lambda, \kappa, \mathbb{P}\right)$.

Empirical Sparse Optimization: Some Limit Theory VI

- Allows for an inferential procedure based on fast subsampling, avoiding the costly use of the Forward Selection Algorithm inside the subsamples.
- $\kappa_{z,T}$ is the solution of $\inf_{\text{csupp}(\kappa)\leq q} \frac{1}{T}\sum_{t=0}^{T}\left(z_t - \sum_{i=0}^{p}\kappa_i X_{t,i}\right)_+$ over $\mathcal{L}_{p,q}$.
- $\Gamma^\star$ is the subset of Γ that contains the triplets at which some accumulation point of $\kappa_{z,T}$ appears.
- For $0 < b_T \leq T$, consider the sub-samples $(X_j)_{j=t,\ldots t+b_T-1}$ for all $t = 1, 2, \ldots, T - b_T + 1$.
- For $\alpha \in (0,1)$, $\text{qu}_{T,B_T}(1-\alpha)$ is the $1-\alpha$ quantile of the empirical distribution of $\left(\sqrt{b_T}\left(\sup_{Z\times\Lambda_p} D\left(z, \kappa_{z,T}, \lambda, \mathbb{P}_{t,b_T}\right) - M^{\text{FS}}\left(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}_T, q\ln^{\epsilon} T\right)\right)\right)_{t=1,\ldots,T-b_T+1}$, where $\mathbb{P}_{t,b_T}$ is the empirical distribution of $(X_j)_{j=t,\ldots,t+b_T-1}$.

Empirical Sparse Optimization: Some Limit Theory VII

Proposition (Fast subsampling KS-type test)

Suppose that: (Condition ND) for the given q, $\Gamma^{\star}$ contains at least one non-trivial triplet. Under the premises of the previous theorem, if $b_T \to \infty$, $\frac{b_T}{T} \to 0$ and $\alpha < \frac{1}{2}$, then the test for the null hypothesis of q-sparse SSD spanning, based on the rejection rule

$$\sqrt{T} M^{\mathrm{FS}} \left(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}_T, q \ln^{\epsilon} T \right) > \mathrm{qu}_{T,B_T} \left(1 - \alpha \right),$$

is asymptotically conservative and consistent.

- Non-triviality occurs when there is downside activation: the portfolio difference loads non-trivially on returns in a non-negligible set of lower-tail states.

Empirical Sparse Optimization: Some Limit Theory VIII

Proposition (Fast subsampling CI)

Under the previous framework:

$$\liminf_{T\to\infty} \mathbb{P}\left[M\left(\Lambda_\infty, \mathcal{L}_{\infty,q}, \mathbb{P}\right) \in \left(Z_{M^{\mathrm{FS}}}(q), M^{\mathrm{FS}}\left(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}_T, q\left(\ln T\right)^{\epsilon}\right) + \frac{q_{T,B_T}\left(1-\alpha\right)}{\sqrt{T}}\right)\right] \geq 1-\alpha, \tag{1}$$

where $Z_{M^{\mathrm{FS}}}(q) := \max(0, M^{\mathrm{FS}}\left(\Lambda, \mathcal{L}_{p,q}, \mathbb{P}_T, q\left(\ln T\right)^{\epsilon}\right) - \frac{q_{T,B_T}(1-\alpha)}{\sqrt{T}})$.

A Toy Monte Carlo Experiment

- Temporal iidness, Gaussianity, $p = 50$ assets; A: 5 assets, $\mu_A = 0.3$ and $\sigma_A = 0.15$; B: 5 assets with $\mu_B = 0.15$ and $\sigma_B = 0.1$; all efficient.
- 40 asset returns with $\mu = 0.1$ and $\sigma = 0.5$; $\rho_{i,j} = 0.001$ $i, j = 1, \ldots, p,\ i \neq j$. (10)-SS-SSD spanned by $A \cup B$.

Sample size T	300	500	1000
Case 1: $q = 5$			
Assets selected:			
Average number	4.82	4.88	4.94
St Deviation	0.0135	0.0108	0.0086
Variability of the Loss:			
Average Loss	0.022	0.015	0.009
Standard Error	0.0003	0.0003	0.0002
Case 2: $q = 10$			
Assets selected:			
Average number	9.86	9.91	9.96
St Deviation	0.0116	0.0102	0.0086
Variability of the Loss:			
Average Loss	0.007	0.004	0.001
Standard Error	10^{-4}	10^{-4}	10^{-4}

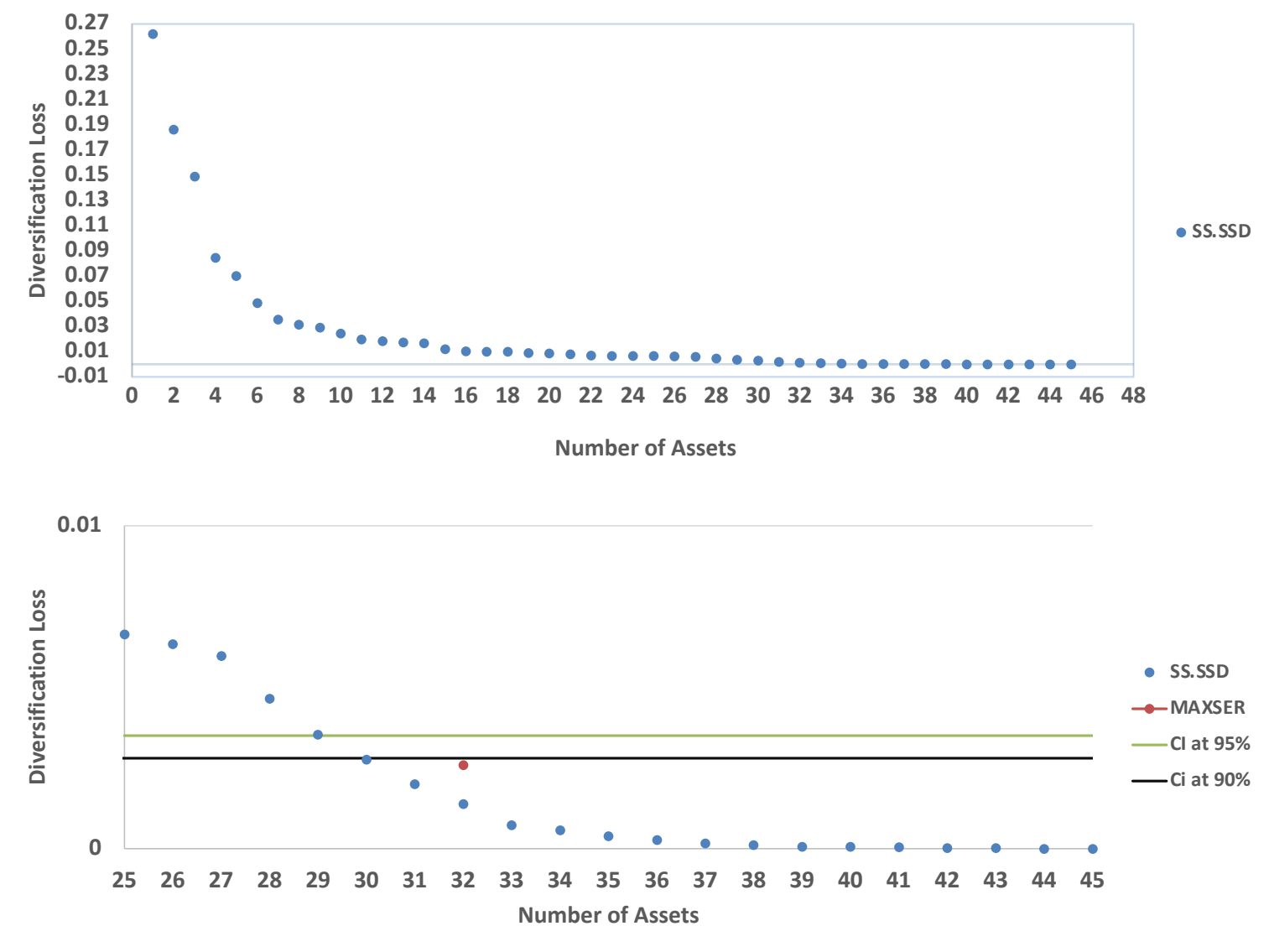
Empirical application

- 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.

Three Optimal Portfolios Compared

- **SS-SSD Portfolio (Sparse SSD-Optimal)**
- **MAXSER Portfolio (Sparse Mean-Variance Optimal):**
  - Based on Ao, Li, and Zheng (2019), using LASSO-type penalization.
  - Optimizes the Sharpe ratio under sparsity constraints.
- **1/N Portfolio (Equally Weighted):**
  - Simple benchmark with equal weights across all p assets.
  - No optimization or sparsity imposed.

The under-diversification loss in the SSD setting



In-Sample Diversification Loss: SSD vs. MAXSER

- For the SSD-based sparse portfolios, the diversification loss $M(q)$ declines rapidly as the support size q increases.
- The loss essentially plateaus around $q \approx 30$–45, indicating limited additional gains from further diversification.
- Relative to MAXSER, the SSD-optimal sparse portfolios exhibit lower downside-risk loss and stronger tail protection.
- Unlike MAXSER, the SSD approach does not rely on covariance matrix estimation or mean–variance efficiency assumptions.
- **Takeaway:** in-sample diversification gains are economically small beyond a moderate number of assets.

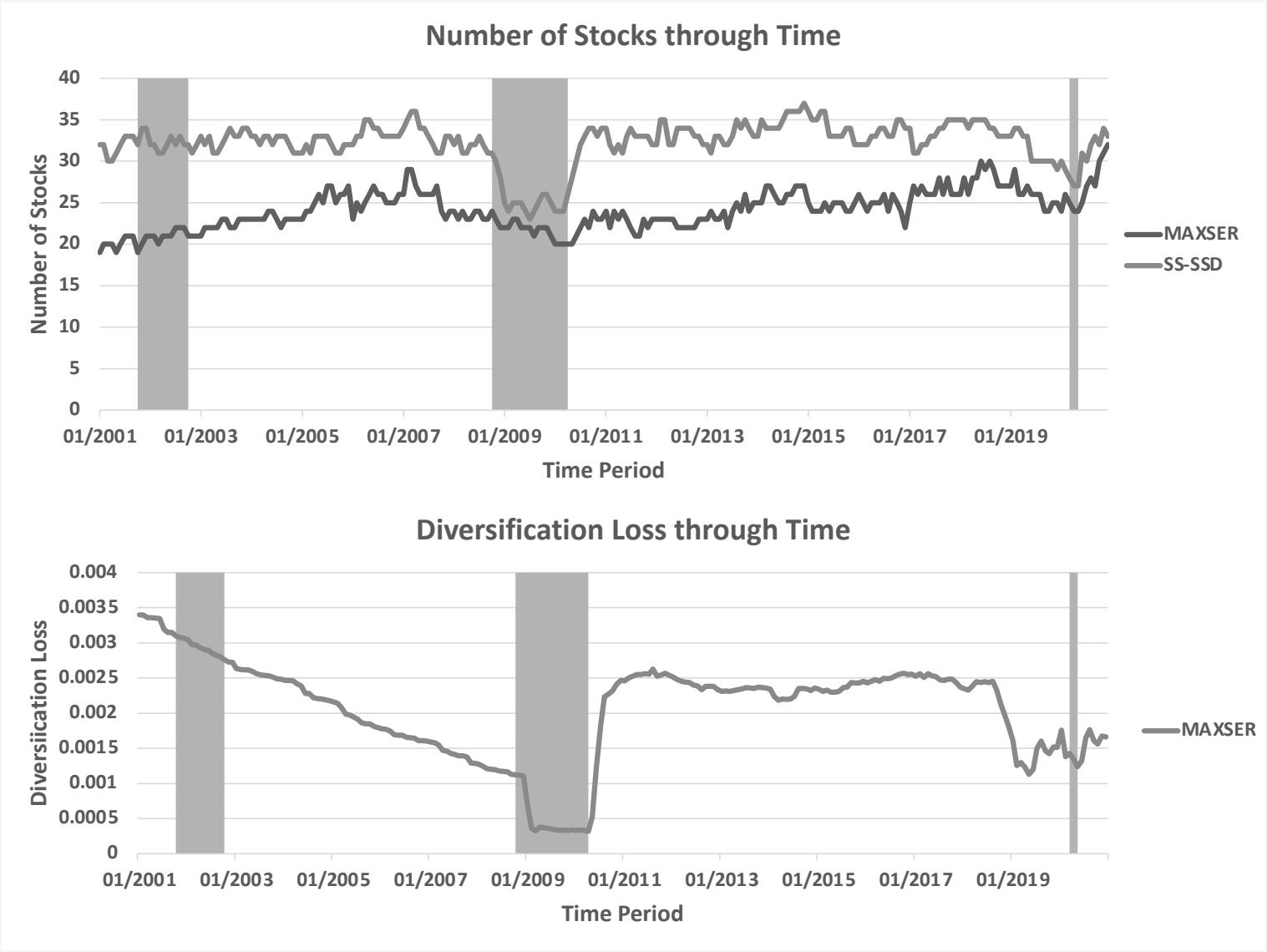
Out-Of-Sample Analysis

- 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.

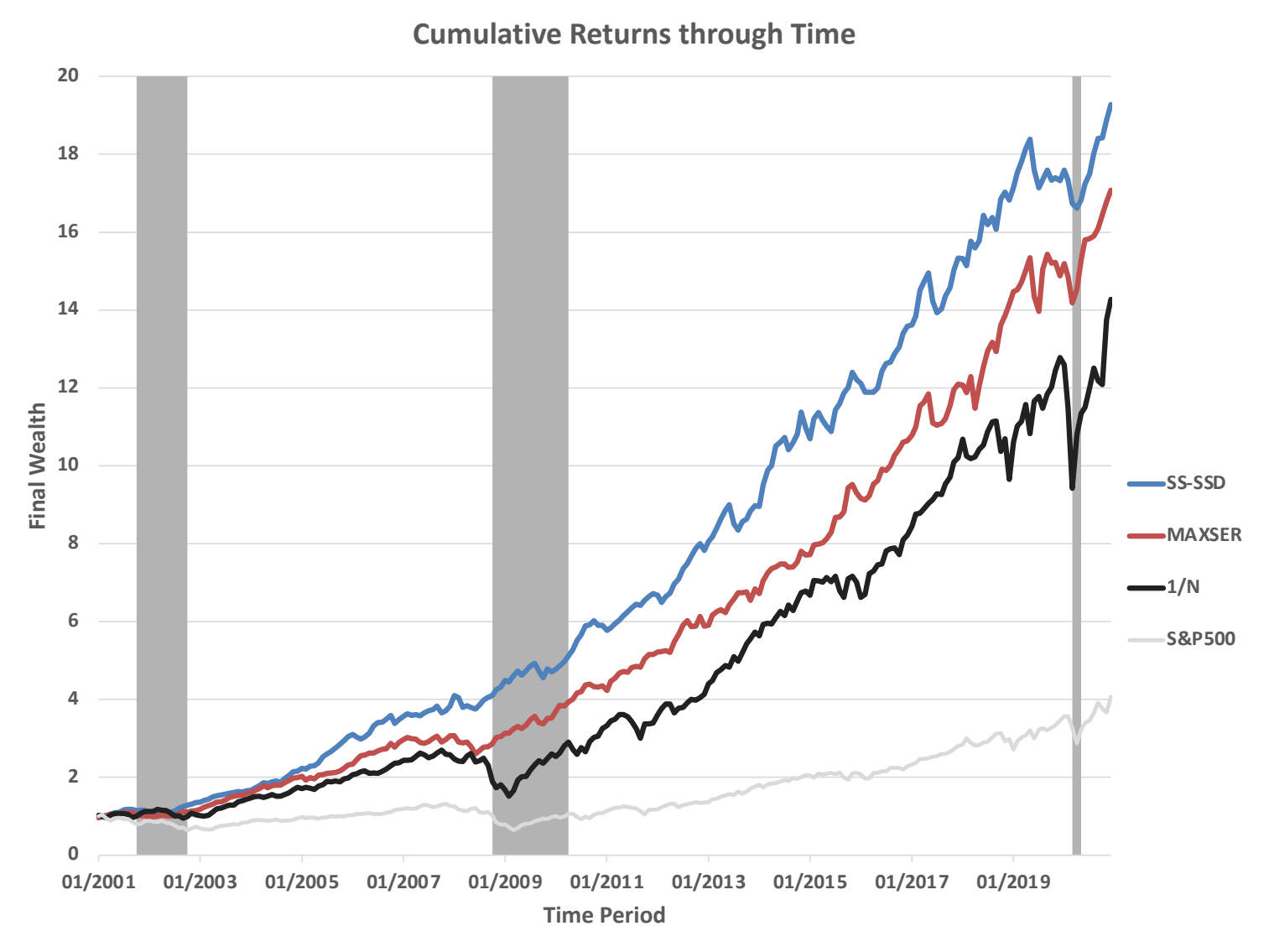
- Each month, portfolios are constructed using the monthly returns during the prior 240 months. The clock is advanced and the realized returns of the optimal portfolios are determined from the actual returns of the various assets. The same procedure is then repeated for the next time period and the ex post realized returns over the period from 01/2001 to 12/2020 (240 months) are computed.

- We again compare the performance of the optimal SS-SSD portfolios with that of the MAXSER portfolios of Ao, Li, and Zheng (2019).

Empirical application: Out-Of-Sample



Out-Of-Sample Analysis



Out-Of-Sample Analysis

Table 1: Out-of-sample performance: risk and performance measures

	MAXSER	SS-SSD	$1/N$
Measures			
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.

Out-Of-Sample Analysis

- We analyze the composition of the SS-SSD and the MAXSER portfolios through time.
- 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.

Out-Of-Sample Analysis: Factor Models

Finally, we investigate which factors explain the returns of the active investors with SSD preferences.

We use CAPM, 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).

These models attempt to explain portfolio returns as linear combinations of common risk factors such as the market, size (SMB), value (HML), profitability (RMW), investment (CMA), and momentum (Mom).

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.

Out of Sample Analysis: Example of Factor Model

Table 2: Fama-French (2016), six factor model

	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
1/N							
Coef.	0.0056	0.9274	0.2262	0.0427	0.1507	0.0877	-0.0511
t-stat	10.009	58.145	9.976	1.697	4.671	2.433	-4.083
p-values	0.0	0.0	0.0	0.0910	0.0	0.0158	0.0

Entries report the coefficients and their respective t-statistics and p-values for the MAXSER portfolio (upper panel), for the SS-SSD portfolio (second panel), and for the $1/N$ portfolio (lower panel). The dataset spans 01/2001-12/2020 for optimal portfolios computed with 240-month windows rolled over one month.

Out-Of-Sample Analysis: Factor Models

- The results indicate that none of the factor models could fully explain the performance of the two strategies, i.e. linear combinations of standard risk factors fail to replicate their returns.

- The intercept a_i is statistically different from zero in all cases, implying the presence of abnormal returns (alphas) not captured by the factors.

- There is no statistically significant factor that explains the returns of the SS-SSD portfolios even if we face a defensive tilt: MKT: <1, SMB: < 0, HML: > 0, indicating that their performance is not well captured by linear risk-factor exposures.

- The results indicate that perhaps nonlinearities, tail-risk effects, or higher-moment features—central to SSD preferences—drive the performance of these portfolios.

Conclusions

- Our new methodology shows that it is empirically plausible to 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.
- The paper focuses on second-order stochastic dominance but could be modified to accommodate other stochastic dominance orders; higher order SD, prospect theory, etc.
- We could then check whether the empirical findings extend in such settings as well.

Conclusions I

- Sparsity with ℓ^p norm constraints?
  - Control over q is lost.
  - ℓ^p norm constraints solutions could provide some robustness guarantees (conservativeness w.r.t. Wasserstein neighborhoods of $\mathbb{P}_T$) via convex duality.
- Reduce computational burden related to sparse optimization:
  - Somehow exploit the embedding of $\mathcal{L}_{p,q}$ to $p-1$ simplex?
- Reduce computational burden related to resampling technologies:
  - Approximate null tails via exponentials depending on topological properties of solutions? (e.g. Euler characteristics)
  - BELR tests with conservative χ^2 based rejection regions via moments selection?
- Improve guarantees using Backward Selection, local search inside Hamming balls?
- Other frameworks? Sparse refinement of partial identification in models based on many moment inequalities with high-dimensional Θ? Entry games models with many firms/markets in IO?