

Portfolio Choice based on Third-degree Stochastic Dominance *

Thierry Post[†] and Miloš Kopa[‡]

March 28, 2016

Abstract

We develop an optimization method for constructing investment portfolios that dominate a given benchmark portfolio in terms of third-degree stochastic dominance. Our approach relies on the properties of the semi-variance function, a refinement of an existing ‘super-convex’ dominance condition and quadratic constrained programming. We apply our method to historical stock market data using an industry momentum strategy. Our enhanced portfolio generates important performance improvements compared with alternatives based on mean-variance dominance and second-degree stochastic dominance. Relative to the CSRP all-share index, our portfolio increases average out-of-sample return by almost seven percentage points per annum without incurring more downside risk, using quarterly rebalancing and without short selling.

Key words: Portfolio choice, Stochastic dominance, Quadratic programming, Enhanced indexing, Industry momentum.

JEL Classification: C61, D81, G11

*We thank Erick Delage, Bogdan Grechuk, Jens Jackwerth and Pim van Vliet for useful comments and suggestions.

[†]T. Post is Professor of Finance at Koç University, Graduate School of Business; 34450 Sarıyer/Istanbul, Turkey; e-mail: thierrypost@hotmail.com. He kindly acknowledges generous research support from the College of Administrative Sciences and Economics and the Graduate School of Business of Koç University.

[‡]M. Kopa is Associate Professor at Charles University Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics; e-mail: kopa@karlin.mff.cuni.cz. He kindly acknowledges the support of the Czech Science Foundation under grant 15-02938S.

1 Introduction

Portfolio optimization based on stochastic dominance (SD) considers the entire probability distribution of investment returns. This approach is theoretically more appealing but analytically more demanding than mainstream mean-variance (MV) analysis, which considers only expected return and standard deviation. Interestingly, modern-day computer hardware and solver software bring SD optimization within reach of practical application.

The original, first-degree stochastic dominance (FSD) criterion by [Quirk and Saposnik \(1962\)](#) relies on the minimal assumption that investors prefer more to less, without restricting the attitude towards risk. Not surprisingly, this criterion can not rank most portfolios and often leads to indecision or sub-optimal solutions. Indeed, most applications of SD to portfolio choice are based on the more powerful second-degree stochastic dominance (SSD) criterion by [Hadar and Russell \(1969\)](#), [Hanoch and Levy \(1969\)](#) and [Rothschild and Stiglitz \(1970\)](#), which assumes that investors are risk averse.

[Shalit and Yitzhaki \(1994\)](#), [Kuosmanen \(2004\)](#), [Roman, Darby-Dowman and Mitra \(2006\)](#), [Scaillet and Topaloglou \(2010\)](#), [Kopa and Post \(2015\)](#), [Armbruster and Delage \(2015\)](#) and [Longarela \(2016\)](#) develop various linear programming (LP) and mixed integer linear programming (MILP) problems to construct an enhanced portfolio that dominates a given benchmark portfolio in terms of SSD. [Clark, Jokung and Kassimatis \(2011\)](#), [Roman, Mitra and Zverovich \(2013\)](#) and [Hodder, Jackwerth and Kolokolova \(2015\)](#) apply SSD optimization to active stock selection and asset allocation.

SSD leads to vast performance improvements over FSD. Nevertheless, SSD often trails MV dominance by allowing for unrealistic preferences over higher-order moment risk. Notably, SSD allows for investors whose risk aversion ($-u''(x)$) increases with wealth and who, consequently, prefer negative skewness to positive skewness. A given pair of portfolios will be deemed incomparable by SSD if these hypothetical ‘skewness haters’ disagree with ‘standard’ investors about the ordering of the portfolios.

To avoid this problem, we develop a portfolio optimization method based on [Whitmore’s \(1970\)](#) third-degree stochastic dominance (TSD). TSD is less restrictive than SSD, because it requires a preference ordering only for ‘skewness lovers’, or those risk averters who exhibit decreasing risk aversion ([Menezes, Geiss and Tressler \(1980\)](#)). This assumption is accepted by financial economists based on compelling theoretical and empirical arguments.

The relaxation of the dominance restriction improves the feasible combinations of return and risk. In particular, TSD is well suited for constructing enhanced portfolios with less downside risk and more upside potential than the benchmark. The SSD criterion ignores these solutions if they are sub-optimal for some skewness haters.

In related work, [Porter, Wart and Ferguson \(1973\)](#), [Bawa \(1975\)](#), [Bawa, Lindenberg and Rafsky \(1979\)](#) and [Bawa *et al.* \(1985\)](#) provide algorithms for TSD comparisons between a finite number of given investment alternatives. [Gotoh and Konno \(2000\)](#) develop mean-risk models that are consistent with TSD. We extend these earlier works by accounting for all (infinitely many) feasible portfolios formed from a discrete set of base assets and for all (infinitely many) relevant utility functions and risk measures.

[Post \(2003, Eq. \(20\)\)](#) and [Post and Versijp \(2007, Section IV\)](#) develop linearizations of the first-order optimality conditions for TSD preferences. We can test whether a given portfolio obeys these conditions using a small linear program. In portfolio management, this approach can be used for in-sample back testing and out-of-sample performance evaluation of a given portfolio. Our study goes one step further by also covering the active portfolio construction phase.

[Section 2](#) presents our analytical framework. Our strategy relies on the properties of the semivariance function, a refinement of an existing ‘super-convexity’ dominance condition and quadratic constrained programming.

[Section 3](#) applies our optimization method to active industry-based asset allocation, following [Hodder, Kolokolova and Jackwerth \(2015\)](#). Since we use an intermediate formation period and a short holding period, the investment strategy in effect exploits known price momentum patterns ([Jegadeesh and Titman \(1993\)](#), [Moskowitz and Grinblatt \(1999\)](#)). Momentum strategies typically use an heuristic approach to portfolio formation. The explicit use of decision theory and optimization seems an interesting addition to the momentum literature.

2 Methodology

2.1 Analytical Strategy

[Kopa and Post \(2015\)](#), [Armbruster and Delage \(2015\)](#) and [Longarela \(2016\)](#) implement SD optimization by searching simultaneously over portfolio weights and piecewise-linear utility functions. Although this approach is exact for SSD, it can provide only an approximation for TSD, because decreasing risk aversion

does not allow for local linearity. High precision for all return levels and admissible utility functions requires an ultra-fine discretization of the return range. Unfortunately, the problem size explodes as one refines the partition, because the number of model variables and constraints grows at a quadratic rate with the number of grid points.

This study uses an alternative strategy based on the formulation of TSD in terms of semivariance, or the second-order lower partial moment (Bawa’s (1975), Fishburn (1977)). At first sight, this route appears analytically challenging, because it requires the evaluation of the semivariance function at a continuum of threshold levels, and, in addition, computing the semivariance involves binary variables to indicate whether portfolio return falls below a given threshold level in a given scenario. Nevertheless, we present tractable solutions to these challenges and the resulting approach is more convenient than searching over piecewise-linear utility functions.

Our strategy employs the ‘super-convex’ TSD (SCTSD) approach of Bawa *et al.* (1985), which requires a reduction of the semivariance at a discrete number of return levels. SCTSD provides a tight sufficient condition for TSD. We modify the original SCTSD approach to get an even tighter sufficient condition based on a piecewise-linear approximation to the semivariance function. In contrast to piecewise-linear approximations of admissible utility functions, our approximation does not require an ultra-fine partition, because the relevant semivariance function is known and simple.

We characterize SCTSD by means of an exact and finite system of linear and convex quadratic constraints. Using this system, we can construct an SCTSD enhanced portfolio by means of convex quadratic constrained programming (QCP). To reduce the computational burden in large-scale applications, we propose an effective method to reduce the number of model variables and constraints using vertex enumeration. Using the reduced problem, we are able to perform large-scale applications in less than a minute of run time using a retail desktop computer and standard solver software.

2.2 Definitions

We consider K distinct base assets with random investment returns $\mathbf{x} \in \mathcal{X}^K$, $\mathcal{X} := [a, b]$, $-\infty < a < b < +\infty$. The portfolio possibility set is represented by the unit simplex $\Lambda := \{\boldsymbol{\lambda} \in \mathbb{R}^K : \boldsymbol{\lambda} \geq \mathbf{0}_K; \boldsymbol{\lambda}'\mathbf{1}_K = 1\}$. Importantly, the base

assets are not restricted to individual securities. In general, the base assets are defined as the most extreme feasible combinations of the individual securities. This formulation allows for general linear weight constraints, including short sales constraints, position limits and restrictions on risk factor loadings. To allow for dynamic intertemporal investment problems, these combinations may be periodically rebalanced based on a conditioning information set. In our application, the base assets are stock portfolios that are formed based on the industry classification of individual stocks.

The returns of the base assets are treated as random variables with a discrete joint probability distribution with T mutually exclusive and exhaustive scenarios with realizations $\mathbf{X}_s := (X_{1,s} \cdots X_{K,s})^T$ and probabilities $p_s := \mathbb{P}[\mathbf{x} = \mathbf{X}_s]$, $s = 1, \dots, T$. The cumulative distribution function for portfolio $\boldsymbol{\lambda} \in \Lambda$ is given by $\mathcal{F}_\lambda(x) := \sum_{t=1}^T p_t D_{\lambda,t}(x)$, where $D_{\lambda,s}(x)$, $s = 1, \dots, T$, is a binary variable that takes a value of one if $\mathbf{X}_s^T \boldsymbol{\lambda} \leq x$ and zero otherwise. Our application analyzes empirical distributions with equally likely, historical scenarios, $p_s = \frac{1}{T}$, $s = 1, \dots, T$.

We evaluate a given and feasible benchmark portfolio $\boldsymbol{\tau} \in \Lambda$. To simplify the notation, we use $y_s := \mathbf{X}_s^T \boldsymbol{\tau}$, $s = 1, \dots, T$, and we assume that the scenarios are ranked in ascending order by the benchmark returns: $y_1 \leq \dots \leq y_T$.

There exist several equivalent formulations of SD criteria. Our analysis uses a common formulation in terms of lower partial moments (Bawa (1975), Fishburn (1977)). For SSD, the first-order lower partial moment (LPM), or expected shortfall, is the relevant risk measure. We use the following definition of the expected shortfall for portfolio $\boldsymbol{\lambda} \in \Lambda$ and threshold return $x \in \mathcal{X}$:

$$\mathcal{E}_\lambda(x) := \sum_{t=1}^T p_t (x - \mathbf{X}_t^T \boldsymbol{\lambda}) D_{\lambda,t}(x). \quad (1)$$

In general, expected shortfall is a continuous, non-negative, non-decreasing and convex function. Under our discrete distribution, $\mathcal{E}_\lambda(x)$ takes a piecewise-linear form with discontinuous increases in its slope at $\mathbf{X}_s^T \boldsymbol{\lambda}$, $s = 1, \dots, T$. The relation between $\mathcal{E}_\lambda(x)$ and $\mathcal{F}_\lambda(x)$ is that $\mathcal{E}_\lambda(x) = \int_a^x \mathcal{F}_\lambda(z) dz$ and $(\partial \mathcal{E}_\lambda(x) / \partial x) = \mathcal{F}_\lambda(x)$.

We use the term ‘expected shortfall’ here to refer to the first-order LPM rather than the Conditional Value at Risk (CVaR). CVaR is the conditional expectation of outcomes below a given percentile; see, for example, Scaillet (2004). These two concepts are closely related. Specifically, $(\mathcal{E}_\lambda(x) / \mathcal{F}_\lambda(x) - x)$

equals the CVaR for the $\mathcal{F}_\lambda(x)$ percentile, or a confidence level of $(1 - \mathcal{F}_\lambda(x))$. Not surprisingly, we can formulate SSD equivalently in terms of the first-order LPM or the CVaR.

Definition 2.2.2: *Portfolio $\lambda \in \Lambda$ dominates the benchmark $\tau \in \Lambda$ by second-degree stochastic dominance (SSD), or $\lambda \succeq_{SSD} \tau$, if*

$$\mathcal{E}_\lambda(y_s) \leq \mathcal{E}_\tau(y_s), \quad s = 1, \dots, T. \quad (2)$$

SSD is a mathematical preorder on the portfolio set Λ : it possesses reflexivity and transitivity but not anti-symmetry, as two distinct portfolios are equivalent if their return vectors are identical ($\mathbf{X}_s^T \lambda = y_s, s = 1, \dots, T$). The economic meaning of the preorder can be explained using the following set of increasing and concave utility functions:

$$\mathcal{U}_2 := \{u \in \mathcal{C}^2(\mathcal{X}) : u'(x) \geq 0; u''(x) \leq 0 \forall x \in \mathcal{X}\}. \quad (3)$$

It is known that $\lambda \succeq_{SSD} \tau$ if and only if $\sum_{t=1}^T p_t u(\mathbf{X}_t^T \lambda) \geq \sum_{t=1}^T p_t u(y_t)$ for all $u \in \mathcal{U}_2$; see [Hadar and Russell \(1969\)](#), [Hanoch and Levy \(1969\)](#) and [Rothschild and Stiglitz \(1970\)](#).

For TSD, the second-order LPM, or semivariance, is the relevant risk measure. We use the following definition of the semivariance for portfolio $\lambda \in \Lambda$ and threshold return $x \in \mathcal{X}$:

$$\mathcal{S}_\lambda^2(x) := \sum_{t=1}^T p_t (x - \mathbf{X}_t^T \lambda)^2 D_{\lambda,t}(x). \quad (4)$$

In general, semivariance is a continuously differentiable, non-negative, non-decreasing and convex function of the threshold return $x \in \mathcal{X}$ (see [Gotoh and Konno \(2000, Thm 3.1\)](#)). For our discrete distribution, $\mathcal{S}_\lambda^2(x)$ takes a piecewise-quadratic form with jumps in its curvature at $\mathbf{X}_s^T \lambda, s = 1, \dots, T$. Also relevant for our analysis is that $\mathcal{S}_\lambda^2(x)$ is a convex function of the portfolio weights λ . Other useful results are $\mathcal{S}_\lambda^2(x) = 2 \int_a^x \mathcal{E}_\lambda(y) dy = 2 \int_a^x \int_a^y \mathcal{F}_\lambda(z) dz dy$ and $(\partial \mathcal{S}_\lambda^2(x) / \partial x) = 2 \mathcal{E}_\lambda(x)$.

Definition 2.2.1: *Portfolio $\lambda \in \Lambda$ dominates the benchmark $\tau \in \Lambda$ by third-degree stochastic dominance (TSD), or $\lambda \succeq_{TSD} \tau$, if*

$$\begin{aligned} \mathcal{S}_\lambda^2(x) &\leq \mathcal{S}_\tau^2(x), \forall x \in \mathcal{X}; \\ \sum_{t=1}^T p_t \mathbf{X}_t^T \boldsymbol{\lambda} &\geq \sum_{t=1}^T p_t y_t. \end{aligned} \tag{5}$$

The economic meaning of TSD can be explained using the following set of utility functions:

$$\mathcal{U}_3 := \{u \in \mathcal{C}^3(\mathcal{X}) : u'(x) \geq 0; u''(x) \leq 0; u'''(x) \geq 0 \forall x \in \mathcal{X}\}. \tag{6}$$

\mathcal{U}_3 imposes the accepted assumption of decreasing risk aversion or skewness love ($u'''(x) \geq 0$) in addition to non-satiation and risk aversion. It is known that $\boldsymbol{\lambda} \succeq_{TSD} \boldsymbol{\tau}$ if and only if $\sum_{t=1}^T p_t u(\mathbf{X}_t^T \boldsymbol{\lambda}) \geq \sum_{t=1}^T p_t u(y_t)$ for all $u \in \mathcal{U}_3$; see [Whitmore's \(1970\)](#).

Since TSD does not require a preference ordering for 'skewness haters', it is easier to establish a TSD relation than an SSD relation. SSD is a sufficient but not necessary condition for TSD: $(\boldsymbol{\lambda} \succeq_{SSD} \boldsymbol{\tau}) \Rightarrow (\boldsymbol{\lambda} \succeq_{TSD} \boldsymbol{\tau})$. This entailment can be derived from $\mathcal{U}_3 \subset \mathcal{U}_2$ or, equivalently, from $\mathcal{S}_\lambda^2(x) = 2 \int_a^x \mathcal{E}_\lambda(y) dy$. It follows that the set of portfolios that dominate the benchmark by TSD is larger than the set of portfolios that dominate it by SSD.

To illustrate the potential improvements, consider gross benchmark returns in three equally likely scenarios ($p_1 = p_2 = p_3 = \frac{1}{3}$) with $y_1 = 0.90$, $y_2 = 1.10$ and $y_3 = 1.30$. A hypothetical example of a TSD enhancement is $\mathbf{X}_1^T \boldsymbol{\lambda} = 0.97$, $\mathbf{X}_2^T \boldsymbol{\lambda} = 1.00$ and $\mathbf{X}_3^T \boldsymbol{\lambda} = z$, $z \geq 1.34$. Every plausible investor will chose this enhancement for its attractive combination of downside risk and upside potential. Nevertheless, the SSD rule does not detect dominance, because the enhanced portfolio increases expected shortfall, without increasing semivariance, for some threshold levels. For example, we have $\mathcal{E}_\lambda(1.1) = 0.077 > 0.067 = \mathcal{E}_\tau(1.1)$, but $\mathcal{S}_\lambda^2(1.1) = 0.009 < 0.013 = \mathcal{S}_\tau^2(1.1)$. MV dominance also does not occur, because the enhanced portfolio has a higher variance than the benchmark for every $z \geq 1.34$.

2.3 Super-Convex TSD

The SSD criterion (2) is formulated using a finite number of threshold levels, because $\mathcal{E}_\lambda(y_s) \leq \mathcal{E}_\tau(y_s)$, $s = 1, \dots, T$, implies $\mathcal{E}_\lambda(x) \leq \mathcal{E}_\tau(x)$ for all $x \in \mathcal{X}$, due to the convex and piecewise-linear shape of expected shortfall. By contrast, the TSD criterion (5) requires the evaluation of $\mathcal{S}_\lambda^2(x)$ at a continuum of threshold levels $x \in \mathcal{X}$.

Following Bawa *et al.* (1985), our analysis uses a tight sufficient condition for TSD that is based on a discretization of the return range \mathcal{X} . Our default specification sets the threshold values equal to the realizations of benchmark: $x = y_s$, $s = 1, \dots, T$, just as in formulation (2) of SSD. The sufficient condition requires minimum levels of slack for the semivariance inequalities (5) at these threshold levels. We pay special attention to the tolerance parameters that control the minimum slack levels.

More specifically, we refine the definition of Bawa *et al.* (1985, Section C.2) in the following way:

Definition 2.3.1: Let $\varepsilon_s \geq 0$, $s = 1, \dots, T$, a series of data-dependent tolerance parameters that are defined as $\varepsilon_1, \varepsilon_2 := 0$ and

$$\varepsilon_s := \left(\frac{\mathcal{S}_\tau^2(y_s)}{\mathcal{S}_\tau^2(y_{s-1}) + 2\mathcal{E}_\tau(y_{s-1})(y_s - y_{s-1})} - 1 \right), \quad s = 3, \dots, T. \quad (7)$$

Portfolio $\lambda \in \Lambda$ dominates the benchmark $\tau \in \Lambda$ by super-convex third-degree stochastic dominance (SCTSD), or $\lambda \succeq_{SCTSD} \tau$, if

$$(1 + \varepsilon_s)\mathcal{S}_\lambda^2(y_s) \leq \mathcal{S}_\tau^2(y_s), \quad s = 1, \dots, T; \quad (8)$$

$$\sum_{t=1}^T p_t \mathbf{X}_t^\top \lambda \geq \sum_{t=1}^T p_t y_t.$$

Some remarks on terminology seem in order. Bawa *et al.* (1985) use the term ‘super-convex’ to indicate that their dominance condition is stronger than Fishburn’s (1974) condition of ‘convex stochastic dominance’. In turn, convex TSD is stronger than pairwise TSD, at least for the analysis of a discrete choice set. For our analysis, which uses a convex portfolio possibility set Λ , convex TSD is equivalent to pairwise TSD by some feasible portfolio $\lambda \in \Lambda$. Nevertheless,

super-convex TSD is stronger than pairwise TSD.

How does our definition differ from the original definition of SCTSD? [Bawa et al. \(1985, Section C.2\)](#) use the same value for all tolerance parameters, or $\varepsilon_s = \varepsilon$, $s = 1, \dots, T$, where the value of ε is selected to ensure $(1 + \varepsilon)\mathcal{S}_\tau^2(y_s) \geq \mathcal{S}_\tau^2(y_{s+1})$, $s = 1, \dots, T - 1$. By contrast, our refinement uses a different value for every tolerance, depending on the expected shortfall and semivariance for the relevant threshold level. Our restrictions give a tighter sufficient condition for TSD and can achieve a given level of accuracy using a rougher partition and hence a smaller problem size.

The following result tightens Theorem 7 of [Bawa et al. \(1985, p. 425\)](#):

Proposition 2.3.1: *If portfolio $\lambda \in \Lambda$ dominates portfolio $\tau \in \Lambda$ by SCTSD, then λ also dominates τ by TSD: $(\lambda \succeq_{SCTSD} \tau) \Rightarrow (\lambda \succeq_{TSD} \tau)$.*

Proof: We need to show that the SCTSD conditions imply $\mathcal{S}_\lambda^2(x) \leq \mathcal{S}_\tau^2(x)$ for all threshold levels $x \in \mathcal{X}$. We provide a separate analysis for various sub-intervals of \mathcal{X} .

First, consider $x \in [a, y_2]$. Since $\varepsilon_1 = 0$, (8) implies $\mathcal{S}_\lambda^2(y_1) \leq \mathcal{S}_\tau^2(y_1) = 0$ and, therefore, $\mathcal{S}_\lambda^2(x) \leq \mathcal{S}_\tau^2(x) = 0$ for all $x \in [a, y_1]$, because $\mathcal{S}_\lambda^2(x)$ and $\mathcal{S}_\tau^2(x)$ are non-decreasing. Since $\varepsilon_2 = 0$, (8) also implies $\mathcal{S}_\lambda^2(y_2) \leq \mathcal{S}_\tau^2(y_2)$. This entailment, together with $\mathcal{S}_\lambda^2(y_1) = 0$, implies $\mathcal{S}_\lambda^2(x) \leq \mathcal{S}_\tau^2(x)$ for all $x \in [y_1, y_2]$. We can demonstrate this with a contradiction. If $\mathcal{S}_\lambda^2(z) > \mathcal{S}_\tau^2(z)$ for some $z \in [y_1, y_2]$ and $\mathcal{S}_\lambda^2(y_1) = 0$, then $\mathcal{E}_\lambda(z) > \mathcal{E}_\tau(z)$, or $(\partial \mathcal{S}_\lambda^2(x)/\partial x)|_{x=z} > (\partial \mathcal{S}_\tau^2(x)/\partial x)|_{x=z}$. It follows that $\mathcal{S}_\lambda^2(x)$ grows faster than $\mathcal{S}_\tau^2(x)$ on $[z, y_2]$, which implies $\mathcal{S}_\lambda^2(x) > \mathcal{S}_\tau^2(x)$ for all $x \in [z, y_2]$ and contradicts $\mathcal{S}_\lambda^2(y_2) \leq \mathcal{S}_\tau^2(y_2)$.

Next, consider $x \in (y_{s-1}, y_s]$ for any $s = 3, \dots, T$. Consider the linear line $t(x) := \mathcal{S}_\tau^2(y_{s-1}) + 2\mathcal{E}_\tau(y_{s-1})(x - y_{s-1})$. The crux of the proof is that, under the SCTSD conditions, $\mathcal{S}_\lambda^2(x) \leq t(x) \leq \mathcal{S}_\tau^2(x)$ for all $x \in (y_{s-1}, y_s]$. Since $(\partial \mathcal{S}_\tau^2(x)/\partial x) = 2\mathcal{E}_\tau(x)$, it follows that $t(x)$ is the tangency line at y_{s-1} . Since $\mathcal{S}_\tau^2(x)$ is convex, the tangency line supports $\mathcal{S}_\tau^2(x)$ from below. Furthermore, given $\varepsilon_{s-1} \geq 0$, (8) implies $\mathcal{S}_\lambda^2(y_{s-1}) \leq \mathcal{S}_\tau^2(y_{s-1}) = t(y_{s-1})$, and, given (7), (8) also implies $\mathcal{S}_\lambda^2(y_s) \leq \mathcal{S}_\tau^2(y_{s-1}) + 2\mathcal{E}_\tau(y_{s-1})(y_s - y_{s-1}) = t(y_s)$. Since $\mathcal{S}_\lambda^2(x)$ is convex, and $t(x)$ is linear, the combined results that $\mathcal{S}_\lambda^2(y_{s-1}) \leq t(y_{s-1})$ and $\mathcal{S}_\lambda^2(y_s) \leq t(y_s)$ imply that $t(x)$ envelops $\mathcal{S}_\lambda^2(x)$ from above on the entire interval.

Finally, consider $x \in (y_T, b]$. $\varepsilon_T \geq 0$ and (8) imply that $\mathcal{S}_\lambda^2(y_T) \leq \mathcal{S}_\tau^2(y_T)$. The SCTSD condition on the means $\sum_{t=1}^T p_t \mathbf{X}_t^\top \lambda \geq \sum_{t=1}^T p_t y_t$ can be rewritten as $(\partial \mathcal{S}_\lambda^2(x)/\partial x) \leq (\partial \mathcal{S}_\tau^2(x)/\partial x)$ for all $x \in (y_T, b]$. Hence, $\mathcal{S}_\lambda^2(y_T) - \mathcal{S}_\tau^2(y_T)$ cannot increase and must remain non-positive on $x \in (y_T, b]$. \square

SCTSD is intended as an approximation of TSD rather than as an alternative SD criterion. In fact, SCTSD is not even a preorder as it does not obey reflexivity, as $\lambda \not\preceq_{SCTSD} \lambda$ for all $\lambda \in \Lambda$. SCTSD does however obey transitivity: $(\lambda_2 \succeq_{SCTSD} \lambda_1 \succeq_{SCTSD} \tau) \Rightarrow (\lambda_2 \succeq_{SCTSD} \tau)$.

As discussed in [Section 2.2](#), $(\lambda \succeq_{SSD} \tau) \Rightarrow (\lambda \succeq_{TSD} \tau)$ due to $\mathcal{U}_3 \subset \mathcal{U}_2$. By contrast, SCTSD is not a necessary condition for SSD: $(\lambda \succeq_{SSD} \tau) \not\Rightarrow (\lambda \succeq_{SCTSD} \tau)$. SSD and SCTSD are two *non-nested* sufficient conditions for TSD. A distinguishing feature of SCTSD is that it approximates TSD as we refine the discretization of the return range \mathcal{X} , whereas SSD is not affected by refinements.

In empirical applications with long time-series, our default specification for the threshold values, $x = y_s, s = 1, \dots, T$, generally yields a tight approximation. However, if the distribution is sparse in the tails, we may further tighten the approximation by locally refining the discretization. To reduce the computational burden, we may also locally lessen the partition, for threshold levels $x = y_s, 1 \ll s \ll T$, in the center of the support, where the data tend to be dense and $\mathcal{S}_\lambda^2(y_s) \approx \mathcal{S}_\lambda^2(y_{s-1})$ for every $\lambda \in \Lambda$.

Figure 1 illustrates our approach using the historical distribution of daily excess returns to the benchmark index used in our application from January 1 through December 31, 2013, our most recent formation period. The solid line gives the semivariance of the benchmark as a function of the threshold level.

The dotted line in Panel A represents the approximation of [Bawa et al. \(1985\)](#) using a partition based on the $T = 252$ daily observations. This approximation in effect multiplies the original semivariance levels by a factor of $(1 + \varepsilon)^{-1} \approx 0.40$. We can establish TSD if the semivariance of an enhanced portfolio lies below this line for all threshold levels. The approximation is poor due to the sparsity of data in the left tail, where the semivariance makes relatively large jumps, leading to a relatively high value of $\varepsilon \approx 1.48$. In order to lower ε , we would have to refine the partition and increase the number of constraints.

The dotted line in Panel B gives the approximation based on our tolerance specification [\(7\)](#). This approach in effect uses a piecewise-linear lower envelope for the semivariance function based on local linear approximation. Using this approach, SCTSD and TSD are hardly distinguishable. A modest deviation occurs in the right tail, where the data is sparse and the curvature of the semivariance function is highest. It is easy to iron out this wrinkle by adding a few additional threshold values between 1.5 and 2.5 percentage points.

[Insert Figure 1 about here.]

Our approach does not require an ultra-fine partition of the return range. Using a rough partition based on just 25 equally spaced grid points yields results that are not materially different from the partition based on the $T = 252$ daily observations. This approach leads to a vast reduction of the computational burden. By contrast, the original formulation of SCTSD does not allow for a rough partition without compromising accuracy. It is also not possible with a rough partition to achieve high accuracy using piecewise linear approximations for all admissible utility functions $u \in \mathcal{U}_3$.

2.4 QCP formulation

Thus far, we have taken a candidate portfolio $\boldsymbol{\lambda} \in \Lambda$ as given. In this case, we can directly check the SCTSD conditions (8) without numerical optimization by simply computing all the relevant semivariance levels $\mathcal{S}_{\boldsymbol{\lambda}}^2(y_s)$, $s = 1, \dots, T$. Searching over the portfolio set for an SCTSD enhanced portfolio however requires numerical optimization. At first sight, the binary variables $D_{\boldsymbol{\lambda},s}(x)$, $s = 1, \dots, T$, in our definition of semivariance (4) seem to require integer programming if the portfolio weights $\boldsymbol{\lambda} \in \Lambda$ are treated as model variables. However, we may avoid integer programming by using the following linearly constrained convex quadratic minimization problem for the semivariance of portfolio $\boldsymbol{\lambda} \in \Lambda$ and threshold $x \in \mathcal{X}$:

$$\begin{aligned} \mathcal{S}_{\boldsymbol{\lambda}}^2(x) &= \min_{\boldsymbol{\theta}} \sum_{t=1}^T p_t \theta_t^2 & (9) \\ \theta_t &\geq x - \mathbf{X}_t^T \boldsymbol{\lambda}, \quad t = 1, \dots, T; \\ \theta_t &\geq 0, \quad t = 1, \dots, T. \end{aligned}$$

The problem is designed such that $\theta_t^* = (x - \mathbf{X}_t^T \boldsymbol{\lambda}) D_{\boldsymbol{\lambda},t}(x)$, $t = 1, \dots, T$, is an optimal solution, which removes the need to use binary variables.

We can apply this quadratic formulation to every threshold $x = y_s$, $s = 1, \dots, T$, in our SCTSD conditions (8). For this purpose, we introduce the model variables $\theta_{s,t}$, $s, t = 1, \dots, T$, to capture the terms $(y_s - \mathbf{X}_t^T \boldsymbol{\lambda}) D_{\boldsymbol{\lambda},t}(y_s)$,

$s, t = 1, \dots, T$. In addition, we treat the portfolio weights $\boldsymbol{\lambda} \in \Lambda$ as model variables, which does not introduce further complications, because the constraints of problem (9) are linear in the portfolio weights.

Combining these insights, we can identify SCTSD enhanced portfolios as solutions to the following system of linear and convex quadratic constraints:

$$\begin{aligned}
(1 + \varepsilon_s) \sum_{t=1}^T p_t \theta_{s,t}^2 &\leq \mathcal{S}_{\boldsymbol{\tau}}^2(y_s), \quad s = 1, \dots, T; \\
-\theta_{s,t} - \mathbf{X}_t^T \boldsymbol{\lambda} &\leq -y_s, \quad s, t = 1, \dots, T; \\
-\sum_{t=1}^T p_t \mathbf{X}_t^T \boldsymbol{\lambda} &\leq -\sum_{t=1}^T p_t y_t; \\
\mathbf{1}_K^T \boldsymbol{\lambda} &= 1; \\
\theta_{s,t} &\geq 0, \quad s, t = 1, \dots, T; \\
\lambda_k &\geq 0, \quad k = 1, \dots, K.
\end{aligned} \tag{10}$$

Any feasible solution $\boldsymbol{\lambda}^*$ to this system dominates the benchmark portfolio $\boldsymbol{\tau}$ by SCTSD (and hence by TSD).

The system involves $(T^2 + K)$ variables and $(T^2 + T + 2)$ constraints, excluding $(T^2 + K)$ non-negativity constraints. The T quadratic inequality constraints are convex, which reflects that semivariance is a convex function of the portfolio weights. The convexity of the constraints in the weights implies that the set of SCTSD enhanced portfolios is convex.

To find an SCTSD enhanced solution, we can develop mathematical programming problems that optimize an objective function given these constraints. Examples of objective functions that are consistent with the TSD criterion are maximizing the expected portfolio return and minimizing the portfolio semivariance for a given threshold level. These objective functions are convex functions of the portfolio weights, and hence we end up with a convex QCP problem.

System (10) ensures that the enhanced portfolio dominates the benchmark. We may ask whether the portfolio is also efficient in the sense that it is not possible to further improve any of the relevant performance criteria without worsening other criteria. The portfolio will be efficient if the objective function assigns a strictly positive weight to increasing the mean and reducing all relevant semivariance levels. [Kopa and Post \(2015, Section 5\)](#) show how to specify criterion weights based on a given utility function.

However, if the objective function assigns a zero weight to some of the criteria, then inefficiency may occur. To avoid this outcome, we may define a secondary objective function that does cover all criteria and solve a second problem that optimizes the secondary objective function given the SCTSD constraints and the optimal value of the primary objective function.

The goal in our application section is to maximize the expected portfolio return. This orientation allows for an easy comparison with heuristic portfolio construction rules and with MV optimization based on the same objective function. Although the enhanced portfolio may not always be fully efficient, the remaining improvement possibilities are negligible in our application: we find no material effects on the in-sample and out-of-sample performance from using a secondary objective function based on semi-variance reductions.

2.5 Problem reduction

Practical applications may involve hundreds of (historical or simulated) scenarios and, in these cases, the raw QCP problem would involve tens or hundreds of thousands of variables and constraints. In order to reduce the memory requirements and run time, problem reduction seems desirable. As discussed in [Section 2.3](#), one way to reduce the problem size is to use less grid points than the number of scenarios. Our refinement of the original SCTSD condition is designed to allow for a relatively rough partition without compromising accuracy. Below, we discuss an alternative reduction method.

The bulk of the problem size stems from the T^2 variables $\theta_{s,t}$, $s, t = 1, \dots, T$, and T^2 constraints $-\theta_{s,t} - \mathbf{X}_t^T \boldsymbol{\lambda} \leq -y_s$, $s, t = 1, \dots, T$, which endogenize the binary variables $D_{\boldsymbol{\lambda},t}(y_s)$, $s, t = 1, \dots, T$. Using a preliminary analysis, we can establish unambiguously whether $D_{\boldsymbol{\lambda}^*,t}(y_s) = 0$ or, alternatively, $D_{\boldsymbol{\lambda}^*,t}(y_s) = 1$, for all solutions $\boldsymbol{\lambda}^*$ to system (10), for many, if not most, scenarios and thresholds. Fixing the values of the binary variables for these scenarios and thresholds leads to a potentially large reduction of the number of variables and constraints.

By construction, a solution portfolio $\boldsymbol{\lambda}^*$ must have a higher mean return and a higher minimum return than the benchmark. Consequently, the solution is an element of the following subset of the portfolio choice set:

$$\tilde{\Lambda} := \left\{ \boldsymbol{\lambda} \in \Lambda : \sum_{t=1}^T p_t \mathbf{X}_t^T \boldsymbol{\lambda} \geq \sum_{t=1}^T p_t y_t; \mathbf{X}_1^T \boldsymbol{\lambda} \geq y_1 \right\}. \quad (11)$$

The constraint $\mathbf{X}_1^T \boldsymbol{\lambda} \geq y_1$ is a necessary condition for $\min_t (\mathbf{X}_t^T \boldsymbol{\lambda}) \geq y_1$. If the base assets are highly positively correlated, as is the case in our application, then $\mathbf{X}_1^T \boldsymbol{\lambda} \approx \min_t (\mathbf{X}_t^T \boldsymbol{\lambda})$ and the necessary condition is tight.

Importantly, $\tilde{\Lambda}$ is a convex polytope and we may therefore reformulate it as the convex hull of its V vertices, or its most extreme portfolios. One practical way to identify the vertices is the vertex enumeration algorithm by [Avis and Fukuda \(1992\)](#). The computational complexity of this algorithm is limited in our case, because Λ has relatively low dimensions (K) and $\tilde{\Lambda}$ includes only two additional constraints. We assume here that the extreme portfolios are known and label their portfolio weights as $\boldsymbol{\nu}_i$, $i = 1, \dots, V$, and their investment returns as $Z_{i,t} := \mathbf{X}_t^T \boldsymbol{\nu}_i$, $t = 1, \dots, T$.

An optimal solution for $\theta_{s,t}$ is $\theta_{s,t}^* = (y_s - \mathbf{X}_t^T \boldsymbol{\lambda}^*) D_{\boldsymbol{\lambda}^*,t}(y_s)$, $s, t = 1, \dots, T$. Since $\boldsymbol{\lambda}^* \in \tilde{\Lambda}$, it follows that $(\min_i Z_{i,t}) \leq \mathbf{X}_t^T \boldsymbol{\lambda}^* \leq (\max_i Z_{i,t})$, $t = 1, \dots, T$. This insight allows us to fix the value of $D_{\boldsymbol{\lambda}^*,t}(y_s)$ when either $(\min_i Z_{i,t}) \geq y_s$ (which implies $D_{\boldsymbol{\lambda}^*,t}(y_s) = 0$) or $(\max_i Z_{i,t}) \leq y_s$ (which implies $D_{\boldsymbol{\lambda}^*,t}(y_s) = 1$).

Define the following three index sets to partition $\mathcal{T} := \{1, \dots, T\}$ for any given $s = 1, \dots, T$:

$$\mathcal{T}_s^- := \left\{ t \in \{1, \dots, T\} : \left(\min_{i=1, \dots, V} Z_{i,t} \right) \geq y_s \right\}; \quad (12)$$

$$\mathcal{T}_s^0 := \left\{ t \in \{1, \dots, T\} : \left(\min_{i=1, \dots, V} Z_{i,t} \right) < y_s < \left(\max_{i=1, \dots, V} Z_{i,t} \right) \right\}; \quad (13)$$

$$\mathcal{T}_s^+ := \left\{ t \in \{1, \dots, T\} : \left(\max_{i=1, \dots, V} Z_{i,t} \right) \leq y_s \right\}. \quad (14)$$

Combining $\theta_{s,t}^* = (y_s - \mathbf{X}_t^T \boldsymbol{\lambda}^*) D_{\boldsymbol{\lambda}^*,t}(y_s)$, $s, t = 1, \dots, T$, and $(\min_i Z_{i,t}) \leq \mathbf{X}_t^T \boldsymbol{\lambda}^* \leq (\max_i Z_{i,t})$, $t = 1, \dots, T$, we find

$$\theta_{s,t}^* = 0 \quad \forall s, t : t \in \mathcal{T}_s^-; \quad (15)$$

$$\theta_{s,t}^* = y_s - \mathbf{X}_t^T \boldsymbol{\lambda}^* \quad \forall s, t : t \in \mathcal{T}_s^+. \quad (16)$$

Importantly, \mathcal{T}_s^- and \mathcal{T}_s^+ , $s = 1, \dots, T$, do not depend on the composition

of λ^* and these sets can therefore be determined prior to the optimization. Substituting the optimal solutions (15)-(16) in (10), we arrive at the following reduced system:

$$\begin{aligned}
(1 + \varepsilon_s) \left(\sum_{t \in \mathcal{T}_s^0} p_t \theta_{s,t}^2 + \sum_{t \in \mathcal{T}_s^+} p_t (y_s - \mathbf{X}_t^\top \boldsymbol{\lambda})^2 \right) &\leq \mathcal{S}_T^2(y_s), \quad s = 1, \dots, T; \quad (17) \\
-\theta_{s,t} - \mathbf{X}_t^\top \boldsymbol{\lambda} &\leq -y_s, \quad s = 1, \dots, T; \quad t \in \mathcal{T}_s^0; \\
-\sum_{t=1}^T p_t \mathbf{X}_t^\top \boldsymbol{\lambda} &\leq -\sum_{t=1}^T p_t y_t; \\
\mathbf{1}_K^\top \boldsymbol{\lambda} &= 1; \\
\theta_{s,t} &\geq 0, \quad s = 1, \dots, T; \quad t \in \mathcal{T}_s^0; \\
\lambda_k &\geq 0, \quad k = 1, \dots, K.
\end{aligned}$$

The number of model variables falls from $(T^2 + K)$ to $(\sum_{s=1}^T \text{card}(\mathcal{T}_s^0) + K)$ and the number of constraints (excluding non-negativity constraints) from $(T^2 + T + 2)$ to $(\sum_{s=1}^T \text{card}(\mathcal{T}_s^0) + T + 2)$. In case of a positive correlation between the base assets, we find $\text{card}(\mathcal{T}_s^0) \ll T$ for the bulk of the threshold levels, $s = 1, \dots, T$, which significantly reduces the problem size.

We can illustrate the magnitude of the potential reduction using our application. The application uses $K = 49$ base assets and, in a typical formation period, $T \geq 250$, historical scenarios. We deliberately start with a fine partition to analyze the computational burden of large-scale applications, the effect of lessening the partition and the effect of the above problem reduction. For $T = 250$, system (10) has more than 62,500 variables and 62,500 constraints. By comparison, the reduced system (17) typically has less than 15,625 variables and 15,625 constraints.

Our computations are performed on a desktop PC with a quad-core Intel i7 processor with 2.93 GHz clock speed and 16GB of RAM and using the IPOPT 3.12.3 solver in GAMS. The median run time (using the reduced system (17)) was about four minutes per formation period. Lessening the partition using 100 equally spaced grid points reduces the run time to less than one minute without loss of accuracy; using 25 grid points reduces the run time to just seconds with only a minimal loss of accuracy.

3 Application

3.1 Industry momentum strategy

We implement an industry momentum strategy in the spirit of [Moskowitz and Grinblatt \(1999\)](#) and [Hodder, Kolokolova and Jackwerth \(2015\)](#) and compare the performance improvements from portfolio optimization based on MV dominance, SSD and SCTSD.

The benchmark is the all-share index from the Center for Research in Security Prices (CRSP) at the Booth School of Business at the University of Chicago, a value-weighted average of common stocks listed on the NYSE, AMEX and NASDAQ stock exchanges. The base assets are 49 value-weighted stock portfolios that are formed by grouping individual stocks based on their four-digit Standard Industrial Classification (SIC) codes ($K = 49$).

Since the base assets are diversified industry portfolios, we do not allow for concentrated positions in individual stocks. In addition, the analysis does not allow for short sales, because the base assets include only long positions in individual stocks. Our strategy can be implemented at lower transactions costs than a typical stock-level long-short strategy. One cost-effective way to implement our strategy would be to buy exchange-traded funds (ETFs) that track specific sector indices.

The joint return distribution is estimated using the empirical distribution in a moving window of historical returns. Our data set consists of daily excess returns from January 3, 1927, through December 31, 2014. We analyze returns in excess of the daily yield to the one-month US government bond index. The nominal returns are from Kenneth French' online data library and the Treasury yields from Ibbotson and Associates. At the start of every quarter from 1928Q1 through 2014Q4, we form four different enhanced portfolios based on the excess returns in a trailing 12-month window. The typical window includes more than 250 trading days ($T \geq 250$).

The first enhanced portfolio is based on a heuristic rule. It is an equal-weighted combination of the 15 industries with the highest average return among the 49 industries. This portfolio captures a large part of the industry momentum effect by simply buying past winner industries in equal proportion. In addition, it is well-diversified and hence will show a comparable risk level as the benchmark. The other three enhanced portfolios are constructed through optimization. The objective is to maximize the mean subject to the restriction that

the enhanced portfolio dominates the benchmark by a given decision criterion (MV dominance, SSD or SCTSD).

The choice for a 12-month formation period and a three-month holding period is motivated by earlier studies of industry momentum. Moskowitz and Grinblatt (1999, Table III) show that buying winner industries is most profitable for an intermediate formation period and a short holding period. Since we do not allow for short selling, we can ignore the fact that selling loser industries works better for a short formation period. Furthermore, industry momentum strategies can use shorter holding periods than stock momentum strategies, because industries, in contrast to individual stocks, do not show short-term price reversals.

An unreported robustness analysis confirms that our specification of the estimation period and holding period is optimal and that our results are driven by the industry momentum effect. Increasing the length of the estimation and/or holding period substantially worsens the out-of-sample performance in our analysis in the same way as in Moskowitz and Grinblatt (1999, Table III). This pattern is also consistent with the performance deterioration that Hodder, Kolokolova and Jackwerth (2015, Section 5.2) report for a long estimation period.

We report in-sample performance and out-of-sample performance for $N = 87$ annual non-overlapping evaluation periods from January 1 through December 31 in every year from 1928 through 2014. For in-sample performance, the evaluation period coincides with the formation period; for out-of-sample performance, the evaluation period consists of four consecutive three-month holding periods, each of which starts at the end of a 12-month formation period. By construction, out-of-sample analysis is not possible for the first year, 1927. For the sake of comparability, our in-sample evaluation also excludes 1927.

Clearly, the in-sample results are based on hindsight and the out-of-sample results are more relevant for portfolio managers. The in-sample results are used here to illustrate the features of our SCTSD optimization method and the differences between the various decision criteria.

3.2 Performance summary

Table I summarizes the performance of the market index ('Bench'), the heuristic portfolio ('Top15') and the three optimized portfolios (MV, SSD and SCTSD).

Also shown is a decomposition of the outperformance (SCTSD-minus-Bench) into components of (Top15-minus-Bench), (MV-minus-Top15), (SSD-minus-MV) and (SCTSD-minus-SSD). The first three columns show the average, across all $N = 87$ formation periods, of the sample mean (\bar{X}), standard deviation (s_X) and skewness (sk_X) of daily returns. The next three columns summarize the annual in-sample returns and the last three columns focus on annual out-of-sample returns.

We measure outperformance using the spread ($X - X_{Bench}$) rather than the residual of a risk factor model. The market betas of the enhanced portfolios are substantially smaller than 1, due to the benchmark risk constraints. In addition, the exposures to the Fama and French (1996) size factor ('SMB') and value factor ('HML') are small, due to the dynamic nature of our strategy and the diversified nature of the industry portfolios. Indeed, the 'three-factor alpha' of the portfolios is even larger than ($\bar{X} - \bar{X}_{Bench}$). Even the exposures to the Carhart (1997) momentum factor ('MOM') are limited, because our strategy relies on industry-level rather than stock-level momentum and on buying winners rather than selling losers.

The t-statistic $t_{\bar{X}} = \bar{X}/(s_X/\sqrt{N})$ is included to measure the level of statistical significance. We may compute the Sharpe ratio by dividing the t-statistic of ($X - X_{Bond}$) by a factor of \sqrt{N} . Similarly, we may compute the information ratio by dividing the t-statistic of ($X - X_{Bench}$) by \sqrt{N} . The usual interpretation of these ratios however does not apply here, as it is not possible to 'scale' the enhanced portfolio without violating the constraints on short sales and benchmark risk. In addition, the ratios do not penalize negative skewness and reward positive skewness.

A more meaningful risk-adjusted performance measure is the certainty equivalent (CE) for a representative utility function. There exist convincing theoretical and empirical arguments to assume that relative risk aversion (RRA) for the average investor is approximately constant and close to the value of one (Meyer and Meyer (2005)). Hence, we report the CE for a logarithmic utility function: $CE_{\lambda} := \exp\left(\sum_{t=1}^T p_t \ln\left(1 + \mathbf{X}_t^T \boldsymbol{\lambda}\right)\right) - 1$.

The average annual excess return to the benchmark is 8.16% in our sample period. The negative skewness of daily returns reflects elevated correlation between stocks during market downswings. Skewness lovers will dislike this unintended side-effect of broad diversification.

The Top15 portfolio outperforms the benchmark by 21.00% per annum in the formation period and by 4.50% in the evaluation period. Further perfor-

mance improvements can be achieved by assigning higher weights to the best performing industries. It is however not possible to form an equal-weighted combination of a smaller number of industries without exceeding the benchmark risk levels. The optimization methods address this issue using explicit benchmark risk constraints.

The MV approach requires that the enhanced portfolio does not exceed the variance of the benchmark. Relative to the Top15 strategy, the optimal solution increases the annual mean by 12.37% in-sample and 1.88% out-of-sample. The improvements are even larger in terms of the CE, because the variance constraint reduces the portfolio risk level. Although the MV portfolio achieves the best return-to-variability ratio, its negative skewness suggests that further return enhancement is possible without exceeding the downside risk levels of the benchmark.

The SSD approach imposes restrictions on expected shortfall rather than variance. Although the return-to-variability deteriorates, the mean and skewness of daily returns improve. These improvements are achieved by a stronger concentration in the best-performing industries. Compared with the MV strategy, the average annual return increases by 1.15% in-sample and 0.24% out-of-sample.

The SCTSD constraints on semivariance are less restrictive than the SSD constraints on expected shortfall. Although the resulting portfolio is often similar to the SSD portfolio, the differences systematically lead to further improvements of the mean and skewness of daily returns. The average annual return increases by an additional 1.04% in-sample and 0.19% out-of-sample. The average spread (SCTSD-MV) is almost twice as high as the average spread (SSD-MV). The relative improvement over SSD is even larger in terms of the t-statistic and CE.

Not surprisingly, the incremental effect of the above strategy refinements is diminishing. The largest improvement stems from simply buying the highest-yielding base assets using a proper formation period and holding period. Optimization with benchmark risk constraints further enhances return for a given risk level. Replacing variance with decision-theoretical risk measures is the icing on the cake. The combined effect of these refinements is that the SCTSD portfolio outperforms the benchmark by 35.56% per annum in-sample and 6.81% out-of-sample.

[Insert Table I about here.]

3.3 Close-up of 2013

Figure 2 illustrates the differences between the three optimized portfolios using the empirical distribution of daily returns in the last formation period, January 1 through December 31, 2013. Panel A shows the achieved reduction in expected shortfall relative to the benchmark ($\mathcal{E}_\tau(x) - \mathcal{E}_\lambda(x)$) for every threshold level x ; similarly, Panel B shows the achieved reduction in semivariance ($\mathcal{S}_\tau^2(x) - \mathcal{S}_\lambda^2(x)$).

In this formation period, the MV portfolio dominates the benchmark by SSD and TSD, as it reduces expected shortfall and semivariance for all x . The portfolio enhances the full-year return by 12.92%. The variance constraint is binding, that is, the portfolio has the same variance level as the benchmark. Due to the negative skewness of the benchmark, this constraint is however not required for managing downside risk.

The SSD portfolio increases the full-year return by a further 4.20%. The restriction on expected shortfall is binding for $x \approx -0.5$. Since SSD is a sufficient condition for TSD, the portfolio also reduces the semivariance for all x . Nevertheless, further return enhancements seem possible for all skewness lovers, because the TSD restrictions are not binding.

Indeed, the SCTSD portfolio raises the full-year return by another 1.01%. The portfolio does not dominate the benchmark by MV dominance, as it has a higher standard deviation. SSD also does not occur, as the portfolio violates the expected shortfall constraint for roughly $x \in [-0.7, 0.2]$. However, the portfolio does reduce the semivariance for all x and hence it dominates the benchmark by TSD.

In the year 2014, all three portfolios, formed using 2013 data, continue to outperform the benchmark. The realized annual return of the MV, SSD and SCTSD portfolios exceeds that of the benchmark by 6.43%, 7.51% and 7.61%, respectively.

[Insert Figure 2 about here.]

To further illustrate the differences between the various decision criteria, we map the enhanced indexing problem to the mean-standard deviation space. In

addition to the CRSP all-share index, we now also consider 20 other benchmark portfolios with different levels of standard deviation: the Minimum-Variance Portfolio (MVP); nine different mixtures of the MVP and the index; the Maximum Mean Portfolio (MMP); nine different mixtures of the MMP and the index. For each of these 21 benchmark portfolios, we construct enhanced portfolios based on MV dominance, SSD and SCTSD. Again, our objective is to maximize the mean subject to the benchmark risk restriction.

Figure 3 plots the means of the three enhanced portfolios (MV, SSD and SCTSD) against the standard deviation of the benchmark. The three portfolios that dominate the index (‘CSRP’) are the same ones as shown in Figure 2. The improvement possibilities are largest for the MVP. This portfolio by construction is MV efficient. However, SD optimization can build portfolios that have substantially higher means and more positive skewness. By contrast, it is not possible to improve the mean of the MMP. For the 20 other portfolios, SCTSD leads to larger improvements than MV and SSD. We also applied FSD using the combinatorial optimization method of [Kopa and Post \(2009\)](#) but found no improvement possibilities for any of the 21 benchmark portfolios.

[Insert Figure 3 about here.]

3.4 Cumulative performance

Figure 4 illustrates the cumulative performance of the three optimization strategies for the entire sample period from 1928 through 2014. Shown is the relative value of each enhanced portfolio, or the ratio of cumulative gross return of the portfolio to the cumulative gross return of the benchmark. Not surprisingly, the return enhancements of six to seven percent per annum translate into exponential value growth over time. In 2014, after 87 years, the MV portfolio is 127.38 times more valuable than the benchmark. The SCTSD portfolio shows the strongest value growth and exceeds the benchmark by a factor of 171.58 in 2014. The benchmark risk of these strategies manifests itself during ‘momentum crashes’, such as the prolonged period of underperformance during the late 1980s and early 1990s. Nevertheless, the SCTSD portfolio leads the other two portfolios during the entire 87-year period. In addition, its relative performance

improves in recent years, after the momentum crash of 1998. The maximum drawdown of the SCTSD portfolio in this sample period is only 36.6%, which is much lower than the maximum drawdown of 68.8% for the benchmark.

[Insert Figure 4 about here.]

4 Conclusions

Our application illustrates the potential improvements from portfolio optimization based on TSD instead of MV dominance or SSD. Benchmark risk restrictions on semivariance allow for a higher mean and skewness than restrictions on variance or expected shortfall. These improvements reflect that concentration in past winner industries creates positive skewness, whereas broad diversification creates negative skewness. The improvements increase the appeal of portfolio construction based on decision theory and optimization compared with heuristic rules.

Despite the pleasing out-of-sample performance in this application, further improvements may come from better estimates for the joint return distribution during the holding period. For example, conditioning on the business cycle and market conditions could help to mitigate crashes of the momentum strategy. Another approach combines the historical returns in the formation period with a prior view about the efficiency of the benchmark index to derive a Bayesian posterior distribution. Our method can be applied to random samples from any given parametric probability distribution or dynamic process. Narrowing the cross-section (K) and lengthening the formation period (T) may also help to reduce estimation error, but this effect has to be balanced against a possible loss of portfolio breadth and signal strength.

Robust optimization methods can reduce the sensitivity to (inevitable) estimation error. The tolerances ε_s , $s = 1, \dots, T$, in (8) seem particularly useful for this purpose. We have tuned these parameters for the exact definition of dominance. Using higher tolerance values can reduce the risk of detecting spurious dominance patterns. By contrast, lower tolerance values can reduce the risk of overlooking dominance relations that are obscured by estimation error.

The latter approach is reminiscent of Almost Stochastic Dominance (Leshno and Levy (2002)), or ASD. Despite the merits of ASD, the current literature offers little guidance for the specification of ‘epsilon’, or the admissible violation area. Post and Kopa (2013, p. 324-325) develop a linearization of ASD and their empirical application casts doubt on the relevance of existing experimental estimates of ‘epsilon’ for portfolio analysis. Without an agreed specification for ‘epsilon’, we prefer to use higher-degree SD rules rather than approximate lower-degree ones.

In a follow-up project, we are developing portfolio optimization based on decreasing absolute risk aversion stochastic dominance (DSD; Vickson’s (1975, 1977) and Bawa (1975)), arguably the most appealing of all SD criteria. TSD is a sufficient but not necessary condition for DSD, suggesting further improvement possibilities for investment performance. For base assets with a limited return range \mathcal{X} and/or with comparable means, the two criteria are often indistinguishable. However, Basso and Pianca (1997) demonstrate that the distinction is important for derivative securities and Post, Fang and Kopa (2015) report important consequences for small-cap stocks.

References

1. Armbruster, B. and E. Delage, 2015, Decision Making under Uncertainty when Preference Information is Incomplete, *Management Science* 61, 111-128.
2. Avis, D. and K. Fukuda, 1992, A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra, *Discrete and Computational Geometry* 8, 295–313.
3. Basso, A. and P. Pianca, 1997, Decreasing Absolute Risk Aversion And Option Pricing, *Management Science* 43, 206–216.
4. Bawa, V.S., 1975, Optimal Rules for Ordering Uncertain Prospects, *Journal of Financial Economics* 2, 95-121.
5. Bawa, V.S., J.N. Bodurtha Jr., M.R. Rao and H.L. Suri, 1985, On Determination of Stochastic Dominance Optimal Sets, *Journal of Finance* 40, 417–431.
6. Bawa, V.S., E.B. Lindenberg and L.C. Ralfsky, 1979, An Algorithm to Determine Stochastic Dominance Admissible Sets, *Management Science* 25, 609-622.

7. Carhart, M.M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57-82.
8. Clark, E., O. Jokung and K. Kassimatis, 2011, Making inefficient market indices efficient, *European Journal of Operational Research* 209, 83–93.
9. Fama, E.F. and K.R. French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55-84.
10. Fishburn, P.C., 1974, Convex Stochastic Dominance with Continuous Distribution Functions, *Journal of Economic Theory* 7, 143-58.
11. Fishburn, P.C., 1977, Mean-Risk Analysis with Risk Associated with Below-Target Returns, *American Economic Review* 67, 116-126.
12. Gotoh, J.-Y. and H. Konno, 2000, Third-degree stochastic dominance and mean-risk analysis, *Management Science* 46, 289-301.
13. Hadar, J. and W.R. Russell, 1969, Rules for Ordering Uncertain Prospects, *American Economic Review* 59, 2–34.
14. Hanoch, G. and H. Levy, 1969, The Efficiency Analysis of Choices Involving Risk, *Review of Economic Studies* 36, 335–346.
15. Hodder, E., J.C. Jackwerth and O. Kolokolova, 2015, Improved Portfolio Choice Using Second-Order Stochastic Dominance, *Review of Finance* 19, 1623-1647.
16. Jegadeesh, N. and S. Titman, 1993, Returns to buying winners and selling losers: implications for stock market efficiency, *Journal of Finance* 48, 65–91.
17. Kopa, M. and Post, Th., 2009, A portfolio optimality test based on the first-order stochastic dominance criterion, *Journal of Financial and Quantitative Analysis* 44, 1103–1124.
18. Kopa, M. and Th. Post, 2015, A General Test for SSD Portfolio Efficiency, *OR Spectrum* 37, 703-734.
19. Kuosmanen, T., 2004, Efficient diversification according to stochastic dominance criteria, *Management Science* 50, 1390-1406.
20. Leshno, M., Levy, H., 2002, Preferred by “all” and preferred by “most” decision makers: Almost stochastic dominance, *Management Science* 48, 1074-1085.
21. Longarela, I.R., 2016, A characterization of the SSD-efficient frontier of portfolio weights by means of a set of mixed-integer linear constraints, *Management Science*, DOI: 10.1287/mnsc.2015.2282.

22. Menezes, C., C. Geiss, and J. Tressler, 1980, Increasing Downside Risk, *American Economic Review* 70, 921-932.
23. Meyer, D. and J. Meyer, 2005, Relative risk aversion: what do we know?, *Journal of Risk and Uncertainty* 31, 243–262.
24. Moskowitz, T.J. and M. Grinblatt, 1999, Do Industries Explain Momentum?, *Journal of Finance* 54, 1249-90.
25. Porter, R.B., J.R. Wart and D.L. Ferguson, 1973, Efficient algorithms for conducting stochastic dominance tests on large numbers of portfolios, *Journal of Financial and Quantitative Analysis* 8, 71-81.
26. Post, Th., 2003, Empirical tests for stochastic dominance efficiency, *Journal of Finance* 58, 1905-1932.
27. Post, Th., and Ph. Versijp, 2007, Multivariate Tests for Stochastic Dominance Efficiency of a Given Portfolio, *Journal of Financial and Quantitative Analysis* 42, 489-515.
28. Post, Th., Y. Fang, and M. Kopa, 2015, Linear Tests for DARA Stochastic Dominance, *Management Science* 61, 1615-1629.
29. Post, Th. and M. Kopa, 2013, General Linear Formulations of Stochastic Dominance Criteria, *European Journal of Operational Research* 230, 321-332.
30. Quirk, J. P. and R. Saposnik, 1962, Admissibility and Measurable Utility Functions, *Review of Economics Studies* 29, 140-146.
31. Roman, D., Darby-Dowman, K. and Mitra, G., 2006, Portfolio construction based on stochastic dominance and target return distributions, *Mathematical Programming* 108, 541-569.
32. Roman, D., G. Mitra and V. Zverovich, 2013, Enhanced indexation based on second-order stochastic dominance, *European Journal of Operational Research* 228, 273-281.
33. Rothschild, M. and J. E. Stiglitz, 1970, Increasing Risk: I. A Definition, *Journal of Economic Theory*, 225-243.
34. Scaillet O., 2004, Nonparametric estimation and sensitivity analysis of expected shortfall, *Mathematical Finance* 14, 115-129.
35. Scaillet, O. and N. Topaloglou, 2010, Testing for stochastic dominance efficiency, *Journal of Business and Economic Statistics* 28, 169–180.
36. Shalit, H. and S. Yitzhaki, 1994, Marginal Conditional Stochastic Dominance, *Management Science* 40, 670–684.

37. Vickson, R.G., 1975, Stochastic Dominance Tests for Decreasing Absolute Risk Aversion. I. Discrete Random Variables, *Management Science* 21, 1438–1446.
38. Vickson, R.G., 1977, Stochastic Dominance Tests for Decreasing Absolute Risk Aversion. II. General random Variables, *Management Science* 23, 478-489.
39. Whitmore, G.A., 1970, Third-degree Stochastic Dominance, *American Economic Review* 60, 457–59.

Table 1: Performance summary

Shown are summary statistics for the investment performance of 5 portfolios. The benchmark ('Bench') is the CRSP all-share index. The heuristic 'Top15' portfolio is an equal-weighted combination of the 15 industries with the highest average return among the 49 industries. The remaining portfolios are formed by maximizing the mean subject to a benchmark risk restriction. The MV portfolio has a lower variance than the benchmark; the SSD portfolio obeys the expected shortfall restrictions (2); the SCTSD portfolio obeys the semivariance restrictions (5). The enhanced portfolios are formed at the beginning of every quarter based on a trailing 12-month window of daily excess returns. We evaluate the 5 portfolios in $N = 87$ non-overlapping periods from January 1 through December 31 in every year from 1928 through 2014. The top rows analyze returns in excess of the Treasury yield ('Bond'). The bottom rows show a decomposition of the outperformance (SCTSD-minus-Bench) into components of (Top15-minus-Bench), (MV-minus-Top15), (SSD-minus-MV) and (SCTSD-minus-SSD). The first three columns show the average, across all $N = 87$ periods, of the sample mean (\bar{X}), standard deviation (s_X) and skewness (sk_X) of daily returns. The next two columns show the average annual return in the formation period together with the associated t-statistic $t_{\bar{X}} = \bar{X}/(s_X/\sqrt{N})$. Also shown is the certainty equivalent (CE) for a logarithmic utility function. The final three columns show the average annual return, t-statistic and CE in the evaluation period.

X	In-sample					Out-of-sample				
	\bar{X}	(s_X)	(sk_X)	\bar{X}	$t_{\bar{X}}$	CE	\bar{X}	$t_{\bar{X}}$	CE	
$X_{Bench} - X_{Bond}$	0.028	0.943	-0.325	8.16	3.70	6.19	8.16	3.70	6.19	
$X_{Top15} - X_{Bond}$	0.091	0.981	-0.434	29.17	8.41	25.90	12.66	4.84	10.10	
$X_{MV} - X_{Bond}$	0.128	0.923	-0.253	41.54	15.21	39.51	14.55	6.33	12.62	
$X_{SSD} - X_{Bond}$	0.131	0.965	-0.019	42.69	15.12	40.50	14.79	6.18	12.71	
$X_{SCTSD} - X_{Bond}$	0.134	0.984	0.032	43.73	15.17	41.45	14.98	6.19	12.86	
$X_{Top15} - X_{Bench}$	0.063	0.352	-0.059	21.00	10.03	19.70	4.50	4.58	3.91	
$X_{MV} - X_{Top15}$	0.038	0.553	0.150	12.37	7.13	13.62	1.88	1.65	2.53	
$X_{SSD} - X_{MV}$	0.003	0.213	0.353	1.15	2.22	0.99	0.24	0.42	0.09	
$X_{SCTSD} - X_{SSD}$	0.003	0.080	0.183	1.04	6.40	0.95	0.19	0.88	0.15	
$X_{SCTSD} - X_{Bench}$	0.106	0.662	0.400	35.56	18.63	35.26	6.81	4.58	6.67	

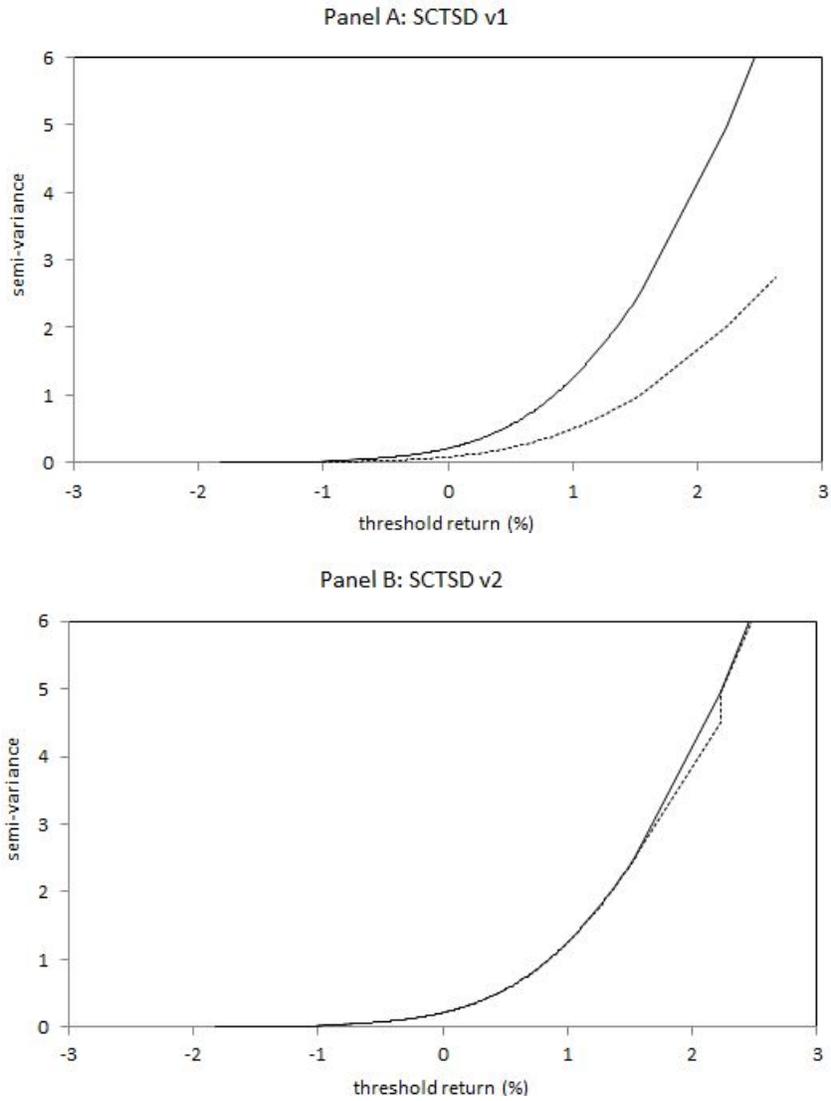


Figure 1: Refinement of the SCTSD condition

Shown are two alternative SCTSD approximations. The solid line in both panels represents the semivariance of the benchmark index as a function of the threshold return in percentage points (%). The figure is based on the daily excess returns in the formation period from January 1 through December 31, 2013 ($T = 252$). The dotted line in Panel A represents the approximation of [Bawa et al. \(1985\)](#) using a partition based on the 252 daily observations. The dotted line in Panel B gives the approximation based on our tolerance specification (7).

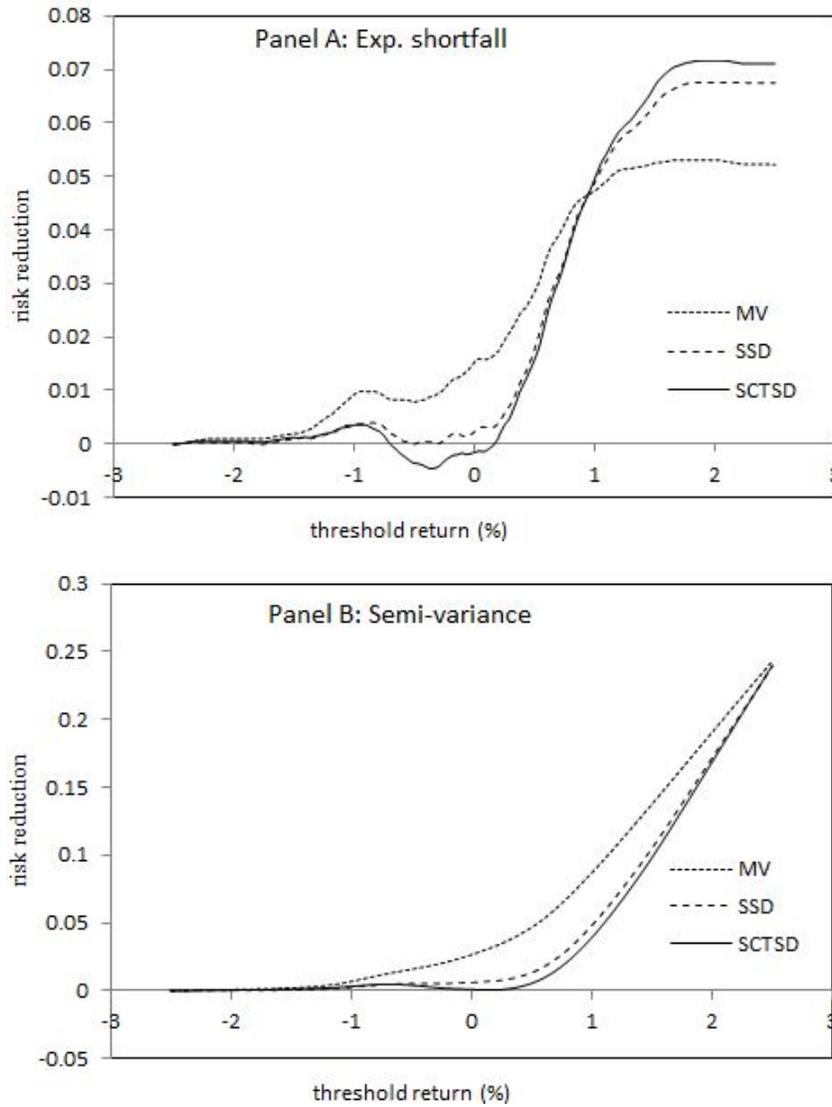


Figure 2: Risk profiles of optimized portfolios

Shown are the risk profiles of the three optimized portfolios (MV, SSD and SCTSD) based on the empirical distribution of daily excess returns in the formation period from January 1 through December 31, 2013. Panel A shows the reduction in expected shortfall ($\mathcal{E}_\tau(x) - \mathcal{E}_\lambda(x)$) for every threshold level x ; similarly, Panel B shows the reduction in semivariance ($\mathcal{S}_\tau^2(x) - \mathcal{S}_\lambda^2(x)$). The returns are in percentage points (%).

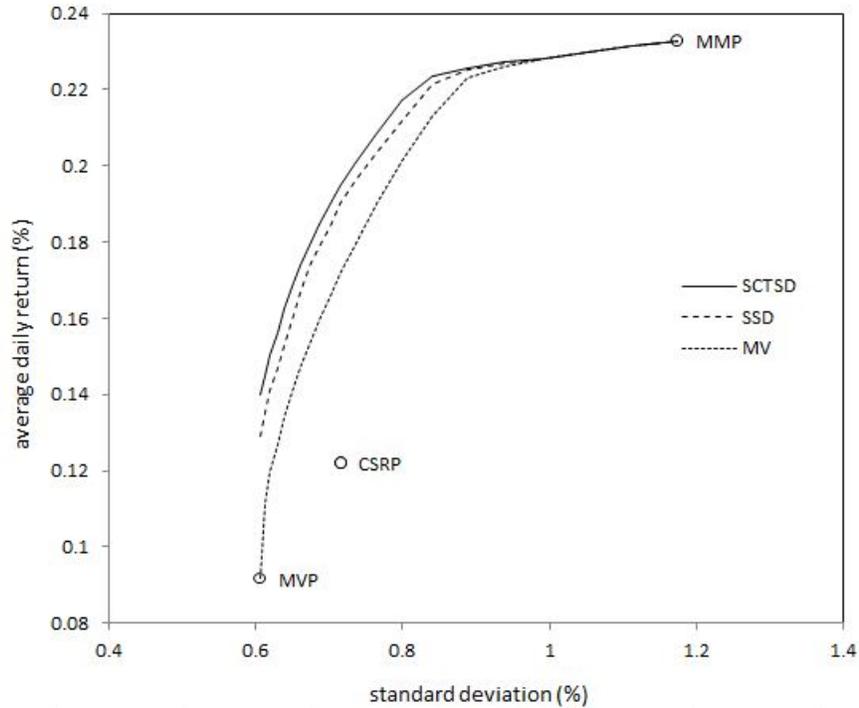


Figure 3: Mean-standard deviation diagram

We analyze 21 benchmark portfolios with different levels of standard deviation: the CRSP all-share index; the Minimum-Variance Portfolio (MVP); nine different mixtures of the MVP and the index; the Maximum Mean Portfolio (MMP); nine different mixtures of the MMP and the index. The weight that the mixed portfolios assign to the index changes with steps of ten percent. For each of the 21 benchmarks, we construct enhanced portfolios based on MV dominance, SSD and SCTSD. In every case, the objective is to maximize the mean subject to the benchmark risk restriction. The figure plots the means of the three enhanced portfolios (MV, SSD and SCTSD) against the standard deviation of the benchmark. The figure is based on the daily excess returns in the formation period from January 1 through December 31, 2013 ($T = 252$).

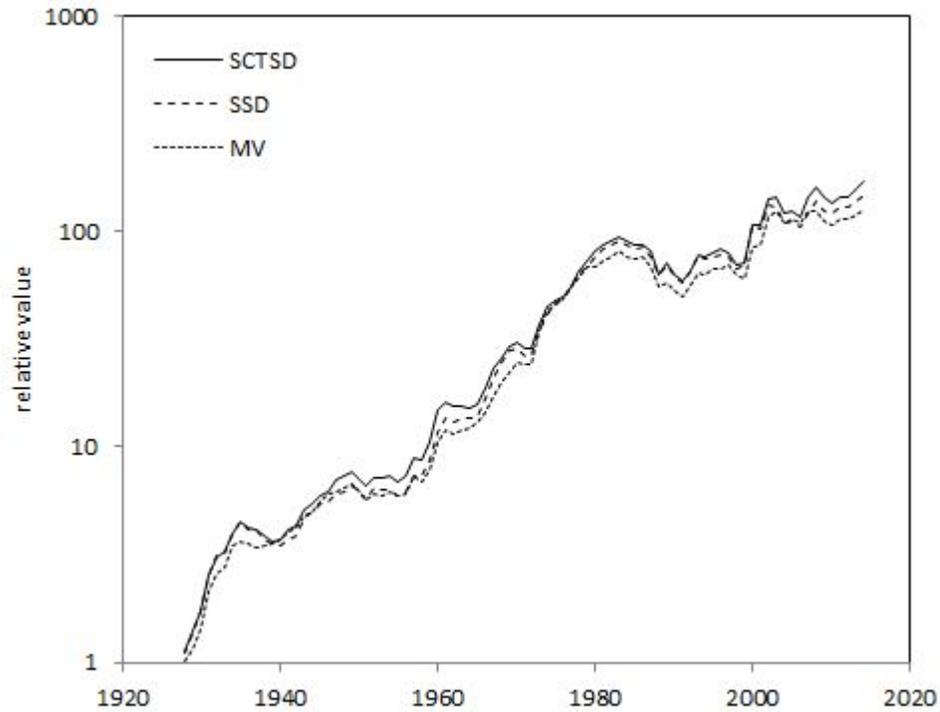


Figure 4: Cumulative performance

Shown is, for each of the three optimized portfolios (MV, SSD, SCTSD), the development of the relative portfolio value over the entire sample period from 1928 through 2014. We measure the relative value as the ratio of cumulative gross return of the enhanced portfolio to the cumulative gross return of the benchmark index. For example, a ratio of 100 in a given year means that the enhanced portfolio has become 100 times more valuable than the benchmark since January 1, 1928. The graph uses a logarithmic scale.