Post, Thierry; Kopa, Miloš

Working Paper

Portfolio choice based on third-degree stochastic dominance, with an application to industry momentum

Working Paper, No. 1527

Provided in Cooperation with:
Koç University - TÜSİAD Economic Research Forum, Istanbul

Suggested Citation: Post, Thierry; Kopa, Miloš (2015) : Portfolio choice based on third-degree stochastic dominance, with an application to industry momentum, Working Paper, No. 1527, Koç University-TÜSİAD Economic Research Forum (ERF), Istanbul

This Version is available at:
https://hdl.handle.net/10419/166722

KOÇ UNIVERSITY-TÜSİAD ECONOMIC RESEARCH FORUM
WORKING PAPER SERIES

PORTFOLIO CHOICE BASED ON THIRD-DEGREE STOCHASTIC DOMINANCE, WITH AN APPLICATION TO INDUSTRY MOMENTUM

Thierry Post
Miloš Kopa

Working Paper 1527
December 2015

KOÇ UNIVERSITY-TÜSİAD ECONOMIC RESEARCH FORUM
Rumelifeneri Yolu 34450 Sarıyer/Istanbul

Portfolio Choice based on Third-degree Stochastic Dominance, With an Application to Industry Momentum*

Thierry Post† & Miloš Kopa‡

November 24, 2015

Abstract

We develop and implement a portfolio optimization method for building investment portfolios that dominate a given benchmark index 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. To reduce the computational burden in large-scale applications, we propose a problem reduction method based on vertex enumeration. 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 benchmark, 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, Semi-variance, Quadratic programming, Enhanced indexing, Industry momentum.

JEL Classification: C61, D81, G11

---

*We would like to thank Bogdan Grechuk, Jens Jackwerth and Pim van Vliet for useful comments and suggestions. Any remaining flaws remain our own.

†Post is Professor of Finance at Koç University, Graduate School of Business; e-mail: thierrypost@hotmail.com

‡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) is theoretically appealing, because it considers the entire probability distribution and uses proper definitions of risk. This approach is analytically more demanding than mainstream mean-variance (MV) analysis, which relies on the convenient but debatable assumption of a Gaussian distribution. However, 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) is overly restrictive as a maxim for investors, because it often leads to indecision or sub-optimal solutions. FSD does not restrict the risk preferences and, consequently, can not compare most portfolios. 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) and Longarela (2015) develop various linear programming (LP) problems to construct an enhanced portfolio that dominates a given benchmark in terms of SSD. Clark, Jokung and Kassimatis (2011), Roman, Mitra and Zverovich (2013) and Hodder, Jackwerth and Kolokolova (2015) report promising results in applications to active stock selection and asset allocation.

Many studies have shown that SSD leads to vast improvements over FSD. Nevertheless, SSD often trails MV dominance, because it allows for unrealistic preferences over higher-order moment risk. Notably, SSD accepts investors whose risk aversion 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 'normal' investors about the ordering of the portfolios.

This study develops an enhanced indexing 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 the '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, the TSD criterion is well suited for constructing enhanced portfolios with less downside risk and more upside potential than the benchmark. The SSD criterion ignores these portfolios 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 portfolios. 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)) develops a linearization of the Karush-Kuhn-Tucker (KKT) conditions for portfolio optimality based on TSD. We can test whether a given portfolio obeys these conditions by solving a small mathematical programming problem. 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.

Following Bawa (1975), we formulate our analysis in terms of lower partial moments. The TSD criterion is analytically challenging because it requires the evaluation of the semi-variance function at a continuum of threshold levels, and, in addition, computing the semi-variance involves binary variables to indicate whether portfolio return falls below a given threshold level in a given scenario.

We obtain a problem of finite dimensions by assuming a discrete state-dependent probability distribution rather than a continuous return distribution. This assumption is not restrictive because practical applications generally use discrete empirical distributions. In addition, many continuous distributions can be approximated accurately using a finite number of random draws.

In addition, we employ the 'super-convex' third-degree stochastic dominance (SCTSD) criterion of Bawa *et al.* (1985), which provides a tight sufficient condition for TSD for discrete return distributions. We modify the original SCTSD criterion to get an even tighter sufficient condition by exploiting the properties of the semi-variance function.

We characterize SCTSD by means of an exact and finite system of linear and convex quadratic constraints. This quadratic system extends the linear formulation of Conditional Value-at-Risk (CVaR) by Rockafellar and Uryasev (2000)

to the case of semi-variance. We can construct an SCTSD enhanced portfolio by solving a convex quadratic constrained programming (QCP) problem. To reduce the computational burden in large-scale applications, we propose an effective method to reduce the number of model variables and model constraints using vertex enumeration. Using this problem reduction, we are able to perform large-scale applications in a just few minutes using a retail desktop computer and standard solver software.

Following Hodder, Kolokolova and Jackwerth (2015), we apply our method to active industry-based asset allocation. 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 SSD and TSD

We consider K distinct base assets with investment returns $\boldsymbol{x} \in \mathcal{X}^K$, $\mathcal{X} := [a, b]$, $-\infty < a < b < +\infty$. 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. To allow for dynamic intertemporal investment problems, these combinations could 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.

Formulated in terms of the individual securities, the portfolio possibility set may take the shape of an arbitrary convex polytope. This formulation allows for general linear weight constraints, including short sales constraints, position limits and restrictions on risk factor loadings. The base assets are the vertices of the polytope. We can therefore formulate the portfolio possibility set as the unit simplex $\Lambda := \left\{\boldsymbol{\lambda} \in \mathbb{R}^K : \boldsymbol{\lambda} \geq \mathbf{0}_K; \boldsymbol{\lambda}'\mathbf{1}_K = 1\right\}$ in the space of the weights of the base assets. This vertex representation of the portfolio possibility set is motivated by the Minkowski-Weyl theorem.

The returns of the base assets are treated as random variables with a state-dependent joint probability distribution with T mutually exclusive and exhaustive scenarios with realizations $\boldsymbol{X}_t := (X_{1,t} \cdots X_{K,t})^{\mathrm{T}}$, $t = 1, \cdots, T$. For our purposes, we can assume that the scenarios are equiprobable, that is, $\mathbb{P}[\boldsymbol{x} =$

$\boldsymbol{X}_t] = \frac{1}{T}$, $t = 1, \cdots, T$. It follows that the cumulative distribution function for portfolio $\boldsymbol{\lambda} \in \Lambda$ is given by $\mathcal{F}_{\boldsymbol{\lambda}}(x) := \frac{1}{T}\sum_{t=1}^{T} \mathbb{I}(\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda} \leq x)$. A heuristic way to deal with unequal probabilities is to include ties, or multiple realizations of the same scenario, in the analysis. Kopa and Post (2015) discuss the computational aspects of this approach.

We evaluate a given and feasible benchmark portfolio $\boldsymbol{\tau} \in \Lambda$. To simplify the notation, we use $\boldsymbol{y}_t := \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\tau}$, $t = 1, \cdots, T$, and we assume that the scenarios are ordered by their ranking with respect to the benchmark: $\boldsymbol{y}_1 \leq \cdots \leq \boldsymbol{y}_T$.

There exist several equivalent formulations of SD criteria. Our analysis uses a common formulation in terms of lower partial moments (Bawa (1975)). For SSD, the first-degree lower partial moment 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 $x \in \mathcal{X}$:

$$\mathcal{E}_{\boldsymbol{\lambda}}(x) := \frac{1}{T}\sum_{t=1}^{T}(x - \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda})\mathbb{I}(\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda} \leq x). \tag{1}$$

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

Definition 2.1: *Portfolio $\boldsymbol{\lambda} \in \Lambda$ dominates the benchmark $\boldsymbol{\tau} \in \Lambda$ by second-degree stochastic dominance (SSD), or $\boldsymbol{\lambda} \succeq_{SSD} \boldsymbol{\tau}$, if*

$$\mathcal{E}_{\boldsymbol{\lambda}}(\boldsymbol{y}_s) \leq \mathcal{E}_{\boldsymbol{\tau}}(\boldsymbol{y}_s),\ s = 1, \cdots, T. \tag{2}$$

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 ($\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda} = \boldsymbol{y}_t$, $t = 1, \cdots, T$). The economic meaning of the preorder can be explained using the following set of increasing and concave utility functions:

$$\mathcal{U}_2 := \left\{u \in \mathcal{C}^2 : \mathcal{X} \to \mathbb{R}; u'(x) \geq 0; u''(x) \leq 0\right\}. \tag{3}$$

It is known that $\boldsymbol{\lambda} \succeq_{SSD} \boldsymbol{\tau}$ if and only if $\frac{1}{T}\sum_{t=1}^{T} u\left(\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda}\right) \geq \frac{1}{T}\sum_{t=1}^{T} u\left(\boldsymbol{y}_t\right)$ 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-degree lower partial moment or semi-variance is the relevant risk measure. We use the following definition of the semi-variance for portfolio $\boldsymbol{\lambda} \in \Lambda$ and threshold return $x \in \mathcal{X}$:

$$\mathcal{S}^2_{\boldsymbol{\lambda}}(x) := \frac{1}{T}\sum_{t=1}^{T}(x - \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda})^2 \mathbb{I}(\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda} \leq x). \tag{4}$$

In general, semi-variance 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, $\mathcal{S}^2_{\boldsymbol{\lambda}}(x)$ takes a piecewise-quadratic form with jumps in its curvature at $\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda}$, $t = 1, \cdots, T$. Also relevant for our analysis is that $\mathcal{S}^2_{\boldsymbol{\lambda}}(x)$ is a convex function of the portfolio weights $\boldsymbol{\lambda}$. Other useful results are $\mathcal{S}^2_{\boldsymbol{\lambda}}(x) = 2\int_a^x \mathcal{E}_{\boldsymbol{\lambda}}(y)dy = 2\int_a^x \int_a^y \mathcal{F}_{\boldsymbol{\lambda}}(z)dzdy$ and $\left(\partial \mathcal{S}^2_{\boldsymbol{\lambda}}(x)/\partial x\right) = 2\mathcal{E}_{\boldsymbol{\lambda}}(x)$.

Definition 2.2: *Portfolio $\boldsymbol{\lambda} \in \Lambda$ dominates the benchmark $\boldsymbol{\tau} \in \Lambda$ by third-degree stochastic dominance (TSD), or $\boldsymbol{\lambda} \succeq_{TSD} \boldsymbol{\tau}$, if*

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

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

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

$\mathcal{U}_3$ 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 $\frac{1}{T}\sum_{t=1}^{T} u\left(\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda}\right) \geq \frac{1}{T}\sum_{t=1}^{T} u\left(\boldsymbol{y}_t\right)$ 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 $\mathcal{S}^2_{\boldsymbol{\lambda}}(x) = 2\int_a^x \mathcal{E}_{\boldsymbol{\lambda}}(y)dy$. 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 of $\boldsymbol{y}_1 = 0.90$, $\boldsymbol{y}_2 = 1.10$ and $\boldsymbol{y}_3 = 1.30$. A hypothetical example of a TSD enhancement is $\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{\lambda} = 0.97$, $\boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{\lambda} = 1.00$ and $\boldsymbol{X}_3^{\mathrm{T}}\boldsymbol{\lambda} = z$, $z \geq 1.34$. 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 semi-variance, for some threshold levels. For example, we have $\mathcal{E}_{\boldsymbol{\lambda}}(1.1) = 0.077 > 0.067 = \mathcal{E}_{\boldsymbol{\tau}}(1.1)$, but $\mathcal{S}^2_{\boldsymbol{\lambda}}(1.1) = 0.009 < 0.013 = \mathcal{S}^2_{\boldsymbol{\tau}}(1.1)$. MV dominance also does not occur, because the enhanced portfolio has a higher variance than the benchmark for every $z \geq 1.34$.

3 Super-Convex TSD

The SSD criterion (2) is formulated using a finite number of threshold levels, because $\mathcal{E}_{\boldsymbol{\lambda}}(\boldsymbol{y}_s) \leq \mathcal{E}_{\boldsymbol{\tau}}(\boldsymbol{y}_s)$, $s = 1, \cdots, T$, implies $\mathcal{E}_{\boldsymbol{\lambda}}(x) \leq \mathcal{E}_{\boldsymbol{\tau}}(x)$ 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 $\mathcal{S}^2_{\boldsymbol{\lambda}}(x)$ 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: $x = \boldsymbol{y}_s$, $s = 1, \cdots, T$, just as in formulation (2) of SSD. The sufficient condition requires minimum levels of slack for the semi-variance 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 3.1: *Let $\varepsilon_s \geq 0$, $s = 1, ..., T$, a series of data-dependent tolerance parameters that are defined as $\varepsilon_1 := -1$, $\varepsilon_2 := 0$ and*

$$\varepsilon_s := \left(\frac{\mathcal{S}^2_{\boldsymbol{\tau}}(\boldsymbol{y}_s)}{\mathcal{S}^2_{\boldsymbol{\tau}}(\boldsymbol{y}_{s-1}) + 2\mathcal{E}_{\boldsymbol{\tau}}(\boldsymbol{y}_{s-1})(\boldsymbol{y}_s - \boldsymbol{y}_{s-1})} - 1 \right), \; s = 3, \cdots, T. \tag{7}$$

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

$$(1+\varepsilon_s)\mathcal{S}^2_{\boldsymbol{\lambda}}(\boldsymbol{y}_s) \leq \mathcal{S}^2_{\boldsymbol{\tau}}(\boldsymbol{y}_s),\ s=1,\cdots,T; \tag{8}$$
$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda} \geq \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{y}_t.$$

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 $\boldsymbol{\lambda} \in \Lambda$. 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 $\varepsilon_s = \varepsilon$, $s = 1, ..., T$, where the value of ε is selected to ensure $(1+\varepsilon)\mathcal{S}^2_{\boldsymbol{\tau}}(\boldsymbol{y}_s) \geq \mathcal{S}^2_{\boldsymbol{\tau}}(\boldsymbol{y}_{s+1})$, $s = 1, ..., T-1$. By contrast, our refinement uses a different value for every tolerance, depending on the expected shortfall and semi-variance 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. One (trivial) manifestation of the improvements is that the semi-variance restriction for $s = 1$ becomes redundant and can be dropped, as $(\varepsilon_1 = -1) \Rightarrow \left((1+\varepsilon_1)\mathcal{S}^2_{\boldsymbol{\lambda}}(\boldsymbol{y}_1) = 0 \leq \mathcal{S}^2_{\boldsymbol{\tau}}(\boldsymbol{y}_1)\right)$.

The following result tightens Theorem 7 of Bawa *et al.* (1985, p. 425):

Proposition 3.1: *If portfolio* $\boldsymbol{\lambda} \in \Lambda$ *dominates portfolio* $\boldsymbol{\tau} \in \Lambda$ *by SCTSD, then* $\boldsymbol{\lambda}$ *also dominates* $\boldsymbol{\tau}$ *by TSD:* $(\boldsymbol{\lambda} \succeq_{SCTSD} \boldsymbol{\tau}) \Rightarrow (\boldsymbol{\lambda} \succeq_{TSD} \boldsymbol{\tau})$.

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

First, consider $x \in [a, \boldsymbol{y}_2]$. We find $\mathcal{S}^2_{\boldsymbol{\tau}}(x) = \begin{cases} 0 & x \leq \boldsymbol{y}_1 \\ (x-\boldsymbol{y}_1)^2 & \boldsymbol{y}_1 < x \leq \boldsymbol{y}_2 \end{cases}$, which has a minimal slope and curvature. If $\mathcal{S}^2_{\boldsymbol{\lambda}}(z) > \mathcal{S}^2_{\boldsymbol{\tau}}(z)$ for some $z \in [a, \boldsymbol{y}_2)$, then the slope and curvature of $\mathcal{S}^2_{\boldsymbol{\lambda}}(x)$ must be at least as high as that of $\mathcal{S}^2_{\boldsymbol{\tau}}(x)$ for all $x \in [z, \boldsymbol{y}_2)$ and hence $\mathcal{S}^2_{\boldsymbol{\lambda}}(\boldsymbol{y}_2) > \mathcal{S}^2_{\boldsymbol{\tau}}(\boldsymbol{y}_2)$. Consequently, the SCTSD restriction $\mathcal{S}^2_{\boldsymbol{\lambda}}(\boldsymbol{y}_2) \leq \mathcal{S}^2_{\boldsymbol{\tau}}(\boldsymbol{y}_2)$ suffices to ensure $\mathcal{S}^2_{\boldsymbol{\lambda}}(x) \leq \mathcal{S}^2_{\boldsymbol{\tau}}(x)$ for all $x \in [a, \boldsymbol{y}_2]$.

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

Finally, consider $x \in (\boldsymbol{y}_T, b]$. $\varepsilon_T \geqslant 0$ and (8) imply that $\mathcal{S}^2_{\boldsymbol{\lambda}}(\boldsymbol{y}_T) \leq \mathcal{S}^2_{\boldsymbol{\tau}}(\boldsymbol{y}_T)$. The SCTSD condition on the means $\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda} \geq \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{y}_t$ can be rewritten as $\left(\partial \mathcal{S}^2_{\boldsymbol{\lambda}}(x)/\partial x\right) \leq \left(\partial \mathcal{S}^2_{\boldsymbol{\tau}}(x)/\partial x\right)$ for all $x \in (\boldsymbol{y}_T, b]$. Hence, $\mathcal{S}^2_{\boldsymbol{\lambda}}(\boldsymbol{y}_T) - \mathcal{S}^2_{\boldsymbol{\tau}}(\boldsymbol{y}_T)$ cannot increase and must remain non-positive on $x \in (\boldsymbol{y}_T, b]$.□

SCTSD is intended as an approximation to TSD rather than as an alternative SD criterion. In fact, SCTSD is not a pre-order as it does not obey reflexivity, as $\boldsymbol{\lambda} \not\succeq_{SCTSD} \boldsymbol{\lambda}$ for all $\boldsymbol{\lambda} \in \Lambda$. SCTSD does however obey transitivity: $(\boldsymbol{\lambda}_1 \succeq_{SCTSD} \boldsymbol{\tau}) \wedge (\boldsymbol{\lambda}_2 \succeq_{SCTSD} \boldsymbol{\lambda}_1) \Rightarrow (\boldsymbol{\lambda}_2 \succeq_{SCTSD} \boldsymbol{\tau})$.

In empirical applications with long time-series, our default specification for the threshold values, $x = \boldsymbol{y}_s$, $s = 1, \cdots, T$, generally yields a tight approximation. It is however possible to further tighten the SCTSD condition by refining the discretization of the return range $\mathcal{X}$. Such refinements are particularly useful if the return distribution is sparse in the tails. To reduce the computational burden, one may also consider lessening the partition, for threshold levels $x = \boldsymbol{y}_s$, $1 \ll s \ll T$, in the center of the support, where the data tend to be dense and $\mathcal{S}^2_{\boldsymbol{\lambda}}(\boldsymbol{y}_s) \approx \mathcal{S}^2_{\boldsymbol{\lambda}}(\boldsymbol{y}_{s-1})$ for every $\boldsymbol{\lambda} \in \Lambda$.

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 semi-variance 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 semi-variance levels by a factor of $(1 + \varepsilon)^{-1} \approx 0.40$. We can establish TSD if the semi-variance 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 semi-variance 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 semi-variance 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 semi-variance 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.]

4 QCP formulation

The binary variables $\mathbb{I}(\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda} \leq x)$, $t = 1, \cdots, T$, in the definition of $\mathcal{S}^2_{\boldsymbol{\lambda}}(x)$ form another complication for optimization. We can however compute $\mathcal{S}^2_{\boldsymbol{\lambda}}(x)$ for a given threshold level $x \in \mathcal{X}$ without binary variables using the following linearly constrained convex quadratic minimization problem in the spirit of the LP problem for CVaR of Rockafellar and Uryasev (2000):

$$\begin{aligned} \mathcal{S}^2_{\boldsymbol{\lambda}}(x) = \min_{\theta} \left(\frac{1}{T} \sum_{t=1}^{T} \theta_t^2 \right) \\ \theta_t \geq x - \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda}, \ t = 1, \cdots, T; \\ \theta_t \geq 0, \ t = 1, \cdots, T. \end{aligned} \tag{9}$$

We can apply this quadratic formulation to every semi-variance level $\mathcal{S}^2_{\boldsymbol{\lambda}}(\boldsymbol{y}_s)$, $s = 1, \cdots, T$, in our SCTSD conditions (8). For this purpose, we introduce the model variables $\theta_{s,t}$, $s = 1, \cdots, T$; $t = 1, \cdots, T$. We can identify an SCTSD enhanced portfolio (as defined in Definition 3.1) as a solution to the following system of linear and quadratic constraints:

$$
\begin{aligned}
(1+\varepsilon_s)\left(\frac{1}{T}\sum_{t=1}^{T}\theta_{s,t}^2\right) &\leq \mathcal{S}_{\boldsymbol{\tau}}^2(\boldsymbol{y}_s),\ s=1,\cdots,T; \qquad (10)\\
-\theta_{s,t}-\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda} &\leq -\boldsymbol{y}_s,\ s=1,\cdots,T;\ t=1,\cdots,T;\\
-\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda} &\leq -\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{y}_t;\\
\mathbf{1}_K^{\mathrm{T}}\boldsymbol{\lambda} &= 1;\\
\theta_{s,t} &\geq 0,\ s=1,\cdots,T;\ t=1,\cdots,T;\\
\lambda_k &\geq 0,\ k=1,\cdots,K.
\end{aligned}
$$

Any feasible solution $\boldsymbol{\lambda}^*$ to system (10) 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, reflecting that semi-variance is a convex function of the portfolio weights. The convexity of the constraints in the weights implies that the set of SCTSD enhanced portfolios, albeit not polyhedral, is convex, as expected.

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 semi-variance for a given threshold level. These objective functions are convex functions of the portfolio weights, and hence we end up with a convex quadratic constrained programming (QCP) problem.

We may ask whether the enhanced portfolio is efficient in the sense that it is not possible to further improve any of the relevant performance criteria without worsening other criteria. The solution will be efficient if the objective function assigns a strictly positive weight to the mean and a strictly negative weight to all relevant semi-variance levels. By contrast, inefficiency may occur if the objective function assigns a zero weight to some of the criteria. In this case, we may define a secondary objective function that covers 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. Kopa and Post (2015, Section 5) show how to specify criterion weights for an arbitrary given utility function.

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. To solve the problem without high-performance computing, some sort of problem reduction seems desirable in order to reduce the memory requirements and run time.

The problem dimensions reflect the number of binary variables, $\mathbb{I}(\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda} \leq \boldsymbol{y}_s)$, $t = 1, \cdots, T$; $s = 1, \cdots, T$, used to compute all relevant semi-variance levels $\mathcal{S}^2_{\boldsymbol{\lambda}}(\boldsymbol{y}_s)$, $s = 1, \cdots, T$, for a given portfolio $\boldsymbol{\lambda} \in \Lambda$. The variables $\theta_{s,t}$, $t = 1, \cdots, T$; $s = 1, \cdots, T$, and constraints $-\theta_{s,t} - \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda} \leq -\boldsymbol{y}_s$, $t = 1, \cdots, T$; $s = 1, \cdots, T$, endogenize the binary variables for all scenarios and thresholds.

However, using a preliminary analysis, we can establish unambiguously whether $\mathbb{I}(\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda}^* \leq \boldsymbol{y}_s) = 0$ or $\mathbb{I}(\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda}^* \leq \boldsymbol{y}_s) = 1$ for many, if not most, scenarios and thresholds for all solutions $\boldsymbol{\lambda}^*$ to system (10). 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 $\boldsymbol{\lambda}^*$ must have a higher mean and a higher minimum than the benchmark. Consequently, the solution is an element of the following convex polytope:

$$\mathrm{K} := \left\{ \boldsymbol{\lambda} \in \Lambda : \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda} \right) \geq \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{y}_t \right) ; \boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{\lambda} \geq \boldsymbol{y}_1 \right\}. \tag{11}$$

We may identify the V vertices of K using the vertex enumeration algorithm by Avis and Fukuda (1992), label the corresponding portfolio weights as $\boldsymbol{\nu}_i$, $i = 1, \cdots, V$, and the investment returns as $Z_{i,t} := \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\nu}_i$, $t = 1, \cdots, T$. Clearly, an optimal solution for $\theta_{s,t}$ is $\theta^*_{s,t} = \max\left(\boldsymbol{y}_s - \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda}^*, 0\right)$, $s = 1, \cdots, T$; $t = 1, \cdots, T$. Since $\boldsymbol{\lambda}^* \in \mathrm{K}$, it follows that $(\min_i Z_{i,t}) \leq \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda}^* \leq (\max_i Z_{i,t})$, $t = 1, \cdots, T$.

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

$$\mathcal{T}_s^- := \left\{ t \in \{1, \cdots, T\} : \left(\min_{i=1,\cdots,V} Z_{i,t} \right) \geq \boldsymbol{y}_s \right\}; \tag{12}$$

$$\mathcal{T}_s^0 := \left\{ t \in \{1, \cdots, T\} : \left(\min_{i=1,\cdots,V} Z_{i,t} \right) < \boldsymbol{y}_s < \left(\max_{i=1,\cdots,V} Z_{i,t} \right) \right\}; \tag{13}$$

$$\mathcal{T}_s^+ := \left\{ t \in \{1, \cdots, T\} : \left(\max_{i=1,\cdots,V} Z_{i,t} \right) \leq \boldsymbol{y}_s \right\}. \tag{14}$$

Combining $\theta_{s,t}^* = \max\left(\boldsymbol{y}_s - \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda}^*, 0\right)$, $s = 1, \cdots, T$; $t = 1, \cdots, T$, and $(\min_i Z_{i,t}) \leq \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda}^* \leq (\max_i Z_{i,t})$, $t = 1, \cdots, T$, we find

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

$$\theta_{s,t}^* = \boldsymbol{y}_s - \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda}^* \; \forall s, t : \; t \in \mathcal{T}_s^+. \tag{16}$$

Importantly, $\mathcal{T}_s^-$ and $\mathcal{T}_s^+$, $s = 1, \cdots, T$, do not depend on the composition of $\boldsymbol{\lambda}^*$ 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:

$$(1+\varepsilon_s)\frac{1}{T}\left(\sum_{t \in \mathcal{T}_s^0} \theta_{s,t}^2 + \sum_{t \in \mathcal{T}_s^+} \left(\boldsymbol{y}_s - \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda} \right)^2 \right) \leq \mathcal{S}_{\boldsymbol{\tau}}^2(\boldsymbol{y}_s), \; s = 1, \cdots, T; \tag{17}$$

$$-\theta_{s,t} - \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda} \leq -\boldsymbol{y}_s, \; s = 1, \cdots, T; \; t \in \mathcal{T}_s^0;$$

$$-\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\lambda} \leq -\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{y}_t;$$

$$\mathbf{1}_K^{\mathrm{T}}\boldsymbol{\lambda} = 1;$$

$$\theta_{s,t} \geq 0, \; s = 1, \cdots, T; \; t \in \mathcal{T}_s^0;$$

$$\lambda_k \geq 0, \; k = 1, \cdots, K.$$

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

The application in the next section illustrates the magnitude of the potential

problem reduction. We construct enhanced portfolios in a situation with $K = 49$ and, in a typical formation period, $T \geq 250$. 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.

6 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 the decision criteria of 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 based on the four-digit Standard Industrial Classification (SIC) code of individual stocks ($K = 49$). Since the base assets include only long positions, we do not allow for short sales. Our strategy will therefore rely on buying recent winner industries rather than selling recent loser industries.

The joint return distribution is forecast 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 2014Q1, 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 recent winner industries in equal proportion. In addition, it is well-diversified and hence will show a comparable risk level as the

benchmark and a relatively high robustness to sampling variation. 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 window 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.

The flip side of using an intermediate formation period is a high sensitivity to sampling variation. We may therefore expect that out-of-sample performance is significantly worse than in-sample performance. Furthermore, using a short holding period generates substantial portfolio turnover and, consequently, transactions costs will consume away part of the outperformance (depending on the available trading facilities). Unfortunately, significant increases of the length of the formation period and holding period reduce the strength of the momentum signal.

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 TSD optimization method and the differences between the various decision criteria.

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 ($\overline{X}$), standard deviation (s_X) and skewness (sk_X) of daily returns. The next two columns show the average annual in-sample return together with the associated t-statistic $t_{\overline{X}} = \overline{X}/(s_X/\sqrt{N})$. The final two columns show the average annual return in the evaluation period and its t-statistic.

The t-statistics are reported here as measures 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.

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 bechmark 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 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 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 performance 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 select the weights by maximizing the mean subject to 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. 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 semi-variance 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 return spread (SCTSD-MV) is almost twice as high on average and less volatile than the spread (SSD-MV).

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.]

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 ($\mathcal{E}_{\boldsymbol{\tau}}(x) - \mathcal{E}_{\boldsymbol{\lambda}}(x)$) for every threshold level x; similarly, Panel B shows the achieved reduction in semi-variance

$(\mathcal{S}^2_{\boldsymbol{\tau}}(x) - \mathcal{S}^2_{\boldsymbol{\lambda}}(x))$.

In this formation period, the MV portfolio dominates the benchmark by SSD and TSD, as it reduces expected shortfall and semi-variance for all x. The portfolio enhances the full-year return by 12.92%. The variance constraint is binding, that is, the MV 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 semi-variance for all x. Nevertheless, further return enhancements seem possible for all skewness lovers, because the TSD restrictions are not binding for the SSD portfolio.

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 semi-variance 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.]

Cumulative performance

Figure 3 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 translates 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 3 about here.]

7 Conclusions

Our application illustrates the potential improvements from portfolio optimization based on TSD instead of MV dominance or SSD. Benchmark risk restrictions on semi-variance allow for a higher mean and skewness than restrictions on variance or expected shortfall. These improvements reflect that concentration in recent 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 forecasts 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 forecast 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) forecast error. The tolerances ε_s, $s = 1, ..., T$, 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. Reversely, lower values can reduce the risk of overlooking dominance relations that are obscured by forecast error.

Future research could focus on portfolio optimization based on decreasing absolute risk aversion stochastic dominance (DARA SD; Vickson's (1975, 1977) and Bawa (1975)), arguably the most appealing of all SD criteria. TSD is a sufficient but not necessary condition for DARA SD, 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. It seems possible to formulate DARA SD optimization as a large convex programming problem, but the implementation currently seems elusive without high-performance computing.

References

1. 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.

2. Basso, A. and P. Pianca, 1997, Decreasing Absolute Risk Aversion And Option Pricing, *Management Science* 43, 206–216.

3. Bawa, V.S., 1975, Optimal Rules for Ordering Uncertain Prospects, *Journal of Financial Economics* 2, 95-121.

4. 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.

5. Bawa, V.S., E.B. Lindenberg and L.C. Rafsky, 1979, An Algorithm to Determine Stochastic Dominance Admissible Sets, *Management Science* 25, 609-622.

6. Carhart, M.M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57-82.

7. Clark, E., O. Jokung and K. Kassimatis, 2011, Making inefficient market indices efficient, *European Journal of Operational Research* 209, 83–93.

8. Fama, E.F. and K.R. French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55-84.

9. Fishburn, P.C., 1974, Convex Stochastic Dominance with Continuous Distribution Functions, *Journal of Economic Theory* 7, 143-58.

10. Gotoh, J.-Y. and H. Konno, 2000, Third-degree stochastic dominance and mean-risk analysis, *Management Science* 46, 289-301.

11. Hadar, J. and W.R. Russell, 1969, Rules for Ordering Uncertain Prospects, *American Economic Review* 59, 2–34.

12. Hanoch, G. and H. Levy, 1969, The Efficiency Analysis of Choices Involving Risk, *Review of Economic Studies* 36, 335–346.

13. Hodder, E., J.C. Jackwerth and O. Kolokolova, 2015, Improved Portfolio Choice Using Second-Order Stochastic Dominance, *Review of Finance* 19, 1623-1647.

14. Jegadeesh, N. and S. Titman, 1993, Returns to buying winners and selling losers: implications for stock market efficiency, *Journal of Finance* 48, 65–91.

15. Kopa, M. and Th. Post, 2015, A General Test for SSD Portfolio Efficiency, *OR Spectrum* 37, 703-734.

16. Kuosmanen, T., 2004, Efficient diversification according to stochastic dominance criteria, *Management Science* 50, 1390-1406.

17. Longarela, I.R., 2015, 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.

18. Menezes, C., C. Geiss, and J. Tressler, 1980, Increasing Downside Risk, *American Economic Review* 70, 921-932.

19. Moskowitz, T.J. and M. Grinblatt, 1999, Do Industries Explain Momentum?, *Journal of Finance* 54, 1249-90.

20. 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.

21. Post, Th., 2003, Empirical tests for stochastic dominance efficiency, *Journal of Finance* 58, 1905-1932.

22. Post, Th., Y. Fang, and M. Kopa, 2015, Linear Tests for DARA Stochastic Dominance, *Management Science* 61, 1615-1629.

23. Quirk, J. P. and R. Saposnik, 1962, Admissibility and Measurable Utility Functions, *Review of Economics Studies* 29, 140-146.

24. Rockafellar, R.T. and S. Uryasev, 2000, Optimization of Conditional Value-at-Risk, *Journal of Risk* 2, 21-41.

25. Roman, D., Darby-Dowman, K. and Mitra, G., 2006, Portfolio construction based on stochastic dominance and target return distributions, *Mathematical Programming* 108, 541-569.

26. Roman, D., G. Mitra and V. Zverovich, 2013, Enhanced indexation based on second-order stochastic dominance, *European Journal of Operational Research* 228, 273-281.

27. Rothschild, M. and J. E. Stiglitz, 1970, Increasing Risk: I. A Definition, *Journal of Economic Theory*, 225-243.

28. Scaillet, O. and N. Topaloglou, 2010, Testing for stochastic dominance efficiency, *Journal of Business and Economic Statistics* 28, 169–180.

29. Shalit, H. and S. Yitzhaki, 1994, Marginal Conditional Stochastic Dominance, *Management Science* 40, 670–684.

30. Vickson, R.G., 1975, Stochastic Dominance Tests for Decreasing Absolute Risk Aversion. I. Discrete Random Variables, *Management Science* 21, 1438–1446.

31. Vickson, R.G., 1977, Stochastic Dominance Tests for Decreasing Absolute Risk Aversion. II. General random Variables, *Management Science* 23, 478-489.

32. Whitmore, G.A., 1970, Third-degree Stochastic Dominance, *American Economic Review* 60, 457–59.

Table I: Performance summary

Shown are summary statistics for the investment performance of 5 portfolios. The benchmark ('Bench') is the CRSP all-share index. The heurististic '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 semi-variance 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 row 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 periods, of the sample mean ($\overline{X}$), 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 $t_{\overline{X}} = \overline{X}/(s_X/\sqrt{N})$. The final two columns show the average annual return in the evaluation period and its t-statistic.

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

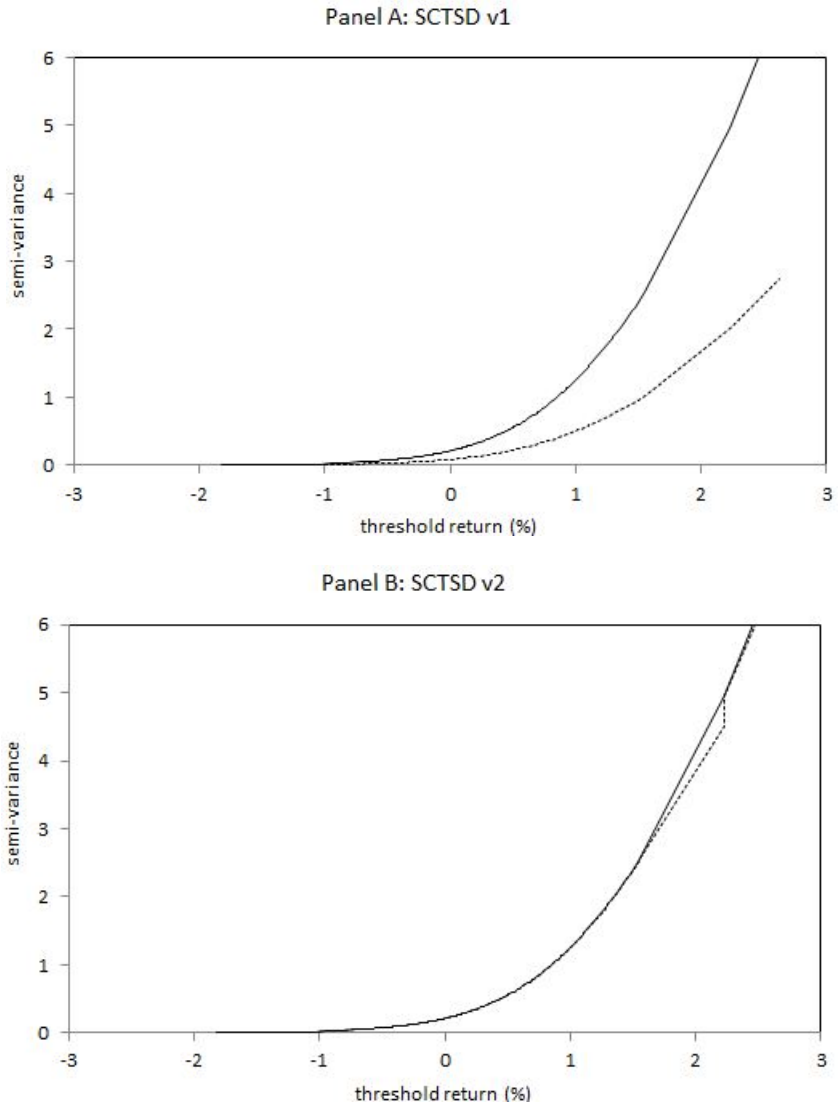

Figure 1: Refinement of the SCTSD condition

Shown are two alternative SCTSD approximations. The solid line in both panels represents the semi-variance 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).

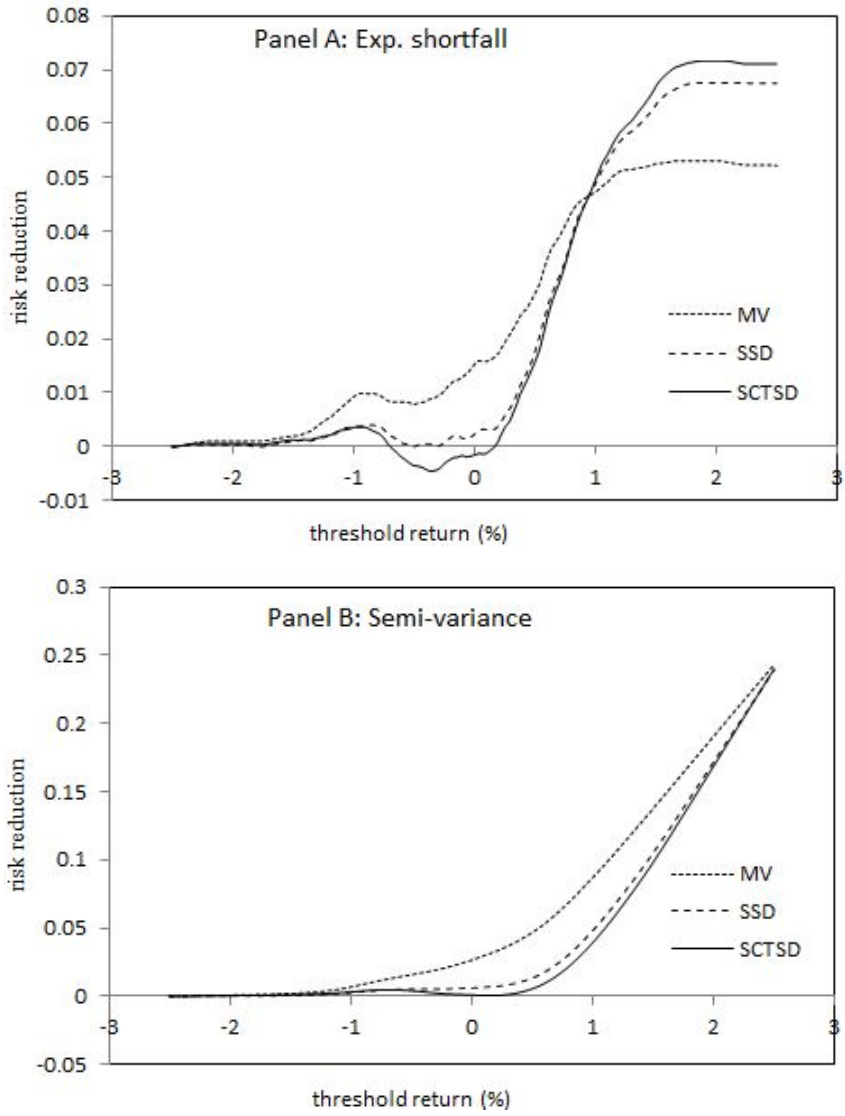

Figure 2: Risk profiles of optimized portfolios

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 ($\mathcal{E}_{\boldsymbol{\tau}}(x) - \mathcal{E}_{\boldsymbol{\lambda}}(x)$) for every threshold level x; similarly, Panel B shows the reduction in semi-variance ($\mathcal{S}^2_{\boldsymbol{\tau}}(x) - \mathcal{S}^2_{\boldsymbol{\lambda}}(x)$). The returns are in percentage points (%)

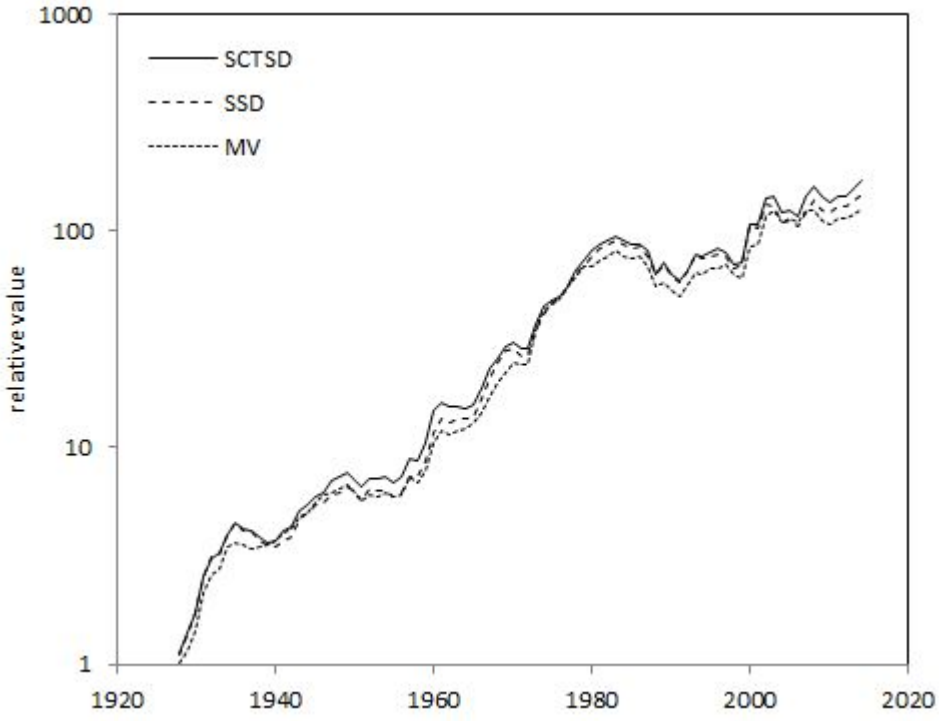

Figure 3: Cumulative performance

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.