Higher-degree Stochastic Dominance Optimality and Efficiency

Yi Fang*and Thierry Post†

March 9, 2017

Abstract

We characterize a range of Stochastic Dominance (SD) relations by means of finite systems of convex inequalities. For 'SD optimality' of degree 1 to 4 and 'SD efficiency' of degree 2 to 5, we obtain exact systems that can be implemented using Linear Programming or Convex Quadratic Programming. For SD optimality of degree five and higher, and SD efficiency of degree six and higher, we obtain necessary conditions. We use separate model variables for the values of the derivatives of all relevant orders at all relevant outcome levels, which allows for preference restrictions beyond the standard sign restrictions. Our systems of inequalities can be interpreted in terms of piecewise polynomial utility functions with a number of pieces that increases with the number of outcomes and the degree of SD. An empirical study analyzes the relevance of higher-order risk preferences for comparing a passive stock market index with actively managed stock portfolios in standard data sets from the empirical asset pricing literature.

Keywords: Decision Analysis, Stochastic Dominance, Expected Utility, Linear Programming, Convex Quadratic Programming.

*Corresponding author. Center for Quantitative Economics, Jilin University, and Business school, Jilin University, Changchun, China; e-mail: danielfang@163.com

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

1 Introduction

Stochastic Dominance (SD) ranks risky prospects based on general regularity conditions for decision making under risk ([QS62], [HR69], [HL69], [RS70], [Whit70]). Recent applications in OR/MS include [LR12], [MXF12], [RMZ13], [PK13], [DK14], [HHM14], [Pod14], [AD15], [EFR16], [Long16], [MSTW16], [PP16] and [PK16].

The classical applications of SD compare a given prospect with a single alternative. More challenging applications involve multiple alternatives. In these cases, the concepts of 'SD optimality' ([Fish74], [BBRS85]) and 'SD efficiency' ([Post03], [DR03], [Kuos04], [PV07], [KP09], [ST10], [Liz12a], [Liz12b], [Post16], [Long16]) apply. In these multivariate applications, a closed-form solution generally does not exist and numerical optimization is required.

Most studies focus on the first three degrees of SD ($N = 1, 2, 3$): first-degree SD (FSD), second-degree SD (SSD) and third-degree SD (TSD). In an ambitious attempt to generalize existing results, [PK13] develop systems of linear inequalities for general Nth degree SD (NSD; $N \geq 1$). With this general formulation, a large class of SD relations can be analyzed using Linear Programming (LP). The relevant LP problems are relatively small and convenient for large-scale applications, simulations and statistical resampling methods.

Despite its merits, the [PK13] approach is not exact but an approximation for SD optimality tests of degree $N \geq 3$ and SD efficiency tests of degree $N \geq 4$. Our study proposes a general revision of [PK13], aiming at stronger operational conditions for higher-degree SD relations. The revision applies to a range of SD relations; we revise even the simple case of pairwise TSD, which arises as a special case of SD optimality with two prospects and $N = 3$.

Our strongest results are obtained for SD optimality of degree $N = 1, 2, 3, 4$ and SD efficiency of degree $N = 2, 3, 4, 5$. For these SD relations, we find finite and exact systems of convex inequalities that can be implemented using LP or Convex Quadratic Programming (CQP). By comparison, the linear systems of [PK13] are exact only for optimality of degree $N = 1, 2$ and efficiency of degree $N = 2, 3$.

For optimality of degree $N \geq 5$ and efficiency of degree $N \geq 6$, our conditions are necessary but not sufficient. We do not consider this an important limitation. The arguments for restricting higher-order derivatives are less compelling than for lower-order derivatives. In addition, these restrictions generally have minimal effects on the flexibility to model the relevant utility levels (for optimality tests) or marginal utility levels (for efficiency tests).

Our analysis introduces model variables for the values of all $(N-1)$ relevant derivatives at all T relevant outcome levels. The additional model variables are not only needed for higher-degree SD relations but can also be used to impose restrictions on the values of the derivatives in addition to the standard restrictions on the signs. This feature is relevant for tests based on Decreasing Absolute Risk Aversion (DARA) SD ([Vick75]), Stochastic Dominance With respect to a Function (SDWRF; [Mey77]), Almost Stochastic Dominance (ASD; [LL02], [LR12], [THS13]) and Standard Stochastic Dominance' (StSD; [Post16]).

One way to interpret our revision is that we use piecewise polynomial functions with a number of pieces that increases with the number of outcomes (T) and the relevant degree of SD (N). This characterization generalizes results by [HS88] and [RS89] on representative utility functions for pairwise comparison based on lower-degree SD rules. Similarly, [CP96, Section 4] derive representative functions of infinite-degree SD, [KP09] and [Post03] deal with the representation of FSD and SSD efficiency and [PFK15, Section 3] with DARA SD optimality and efficiency.

We focus on SD optimality and efficiency tests for a given prospect. The problem of constructing a portfolio which stochastically dominates a given benchmark portfolio ([SY94], [DR03], [Kuos04], [RDM06]) is beyond the scope of this study. However, there exists a close link between these two topics. Notably, [KP15], [AD15] and [Long16] construct SSD efficient portfolios by searching simultaneously over portfolio weights and utility functions using LP. Our results could be used to extend their results to TSD, fourth-degree SD (FOSD) and fifth-order SD (FISD) using CQP.

In an empirical study, we apply a range of portfolio efficiency tests to compare a passive stock market index with actively managed stock portfolios, in standard data sets from the empirical asset pricing literature. Our results show that the estimated pricing errors based on higher-order SD, as well as modifications of SSD based on SDWRF and ASSD, tend to be larger and more significant than standard mean-variance (MV) estimates, as a result of using pricing kernels that exclude arbitrage opportunities and account for systematic skewness. These findings add to the mounting evidence against market portfolio efficiency.

Appendix A presents formal proofs for our lemmas and propositions; Appendix B specifies the LP and CQP problems that we use for our numerical example in Section 7 and empirical application in Section 9.

2 Preliminaries

We use the general framework of [PK13]. Their analysis considers $M \geq 2$ prospects with risky outcomes $x_1, ..., x_M \in \mathcal{D} := [A, B]$, $-\infty < A < B < +\infty$. The outcomes are treated as random variables with a discrete joint probability distribution characterized by R mutually exclusive and exhaustive scenarios with probabilities $p_r > 0$, $r = 1, \cdots, R$.

We use $x_{i,r}$ for the outcome of prospect i in scenario r. We collect all possible outcomes in the joint support $Y := \{y : y = x_{i,r}\ i = 1, ..., M; r = 1, ..., R\}$, rank these values in ascending order, $y_1 \leq \cdots \leq y_S$, and use $p^*_{i;s} := \mathbb{P}[x_i = y_s] = \sum_{r=1}^{R} p_r \mathbb{I}(x_{i,r} = y_s)$, $i = 1, ..., M$; $s = 1, ..., S$.

Decision makers' preferences are described by von Neumann–Morgenstern utility functions. To implement SD of degree $N \geq 1$, we consider the following set of monotonic utility functions:

$$\mathcal{U}_N := \left\{ u \in \mathcal{C}^N : (-1)^{n+1} u^n(x) \geq 0,\ n = 0, \cdots, N \right\}, \tag{1}$$

where $u^0(x) = u(x)$ and $u^n(x) := \partial^n u / \partial x^n$, $n = 1, \cdots, N$.

The economic interpretation of the restrictions on the first two derivatives is well-established: $u^1(x) \geq 0$ amounts to non-satiation and $u^2(x) \leq 0$ means risk aversion. The higher-order derivatives govern the higher-order risk preferences. Notably, $u^3(x) \geq 0$ means 'prudence', or skewness preference, and $u^4(x) \leq 0$ equals 'temperance', or kurtosis aversion. For discussions of the behavioral characterization and consequences of higher-order risk preferences, we refer to [ES06] and [EFR16] and references therein.

The utility set $\mathcal{U}_N$ has two redundant but convenient features. First, the restriction $u(x) \leq 0$ is redundant, because utility analysis is location invariant. This restriction is however convenient because it implies $-u^1(x) \in \mathcal{U}_{N-1}$, which is a useful property in Section 6. Since the below definitions do not require the values of $u^N(x)$, the requirement that the Nth derivative is continuous is also redundant and $\mathcal{U}_N$ is equivalent to

$$\mathcal{U}^*_N := \left\{ u \in \mathcal{C}^{N-1} : (-1)^n \left(u^n(y) - u^n(x) \right) \geq 0,\ n = 0, \cdots, N-1; y \geq x \right\}.$$

The use of $\mathcal{U}_N$ is however convenient to derive Lemma 1 without using subdifferential calculus. However, in Lemma 2 and Section 7, we use $\mathcal{U}_N^*$ to allow for jumps in the Nth derivative.

Definition 1 (Stochastic Dominance). An evaluated prospect x_i, $i = 1, \cdots, M$, is dominated by alternative x_j, $j = 1, \cdots, M$, in terms of NSD, $N \geq 1$, if the former is strictly preferred to the latter for all permissible utility functions $u \in \mathcal{U}_N$:

$$\sum_{r=1}^{R} p_r u(x_{i,r}) < \sum_{r=1}^{R} p_r u(x_{j,r})$$
$$\Leftrightarrow \sum_{s=1}^{S} u(y_s) \left(p_{i,s}^* - p_{j,s}^*\right) < 0. \tag{2}$$

Various applications of SD consider a discrete choice set, $\mathcal{X}_0 := \{x_1, \cdots, x_M\}$, $M \geq 2$. This specification is relevant in welfare economics, where SD is widely applied following [Atkin70], because it is not possible to mix welfare distributions from different countries or periods. Similarly, in health economics, medical treatments are often indivisible and mutually exclusive.

Definition 2 (SD admissibility). An evaluated prospect x_i, $i = 1, \cdots, M$, is admissible in terms of NSD, $N \geq 1$, if it is not dominated by any alternative combination $x \in \mathcal{X}_0$, in terms of NSD.

Algorithms for implementing this concept in an efficient manner were developed in [PWF73] and [BLR79]. The admissibility concept however became obsolete after [BBRS85] developed LP programs to implement a more powerful concept by [Fish74]:

Definition 2' (SD optimality). An evaluated prospect x_i, $i = 1, \cdots, M$, is optimal in terms of NSD, $N \geq 1$, if it is preferred to every alternative $x \in \mathcal{X}_0$ for some permissible utility function $u \in \mathcal{U}_N$:

$$\sum_{r=1}^{R} p_r u(x_{i,r}) \geq \sum_{r=1}^{R} p_r u(x_r) \ \forall x \in \mathcal{X}_0$$
$$\Leftrightarrow \sum_{s=1}^{S} u(y_s) \left(p_{i,s}^* - p_{j,s}^*\right) \geq 0, \ j = 1, \cdots, M. \tag{3}$$

For $M = 2$, the two definitions are equivalent. However, for $M > 2$, Definition 2 is a necessary but not sufficient condition for Definition 2'. Put differently, a prospect can be non-optimal for all permissible utility functions without being dominated by any individual alternative.

In portfolio choice problems, the feasible set generally consists of all convex combinations of the prospects, $\mathcal{X}_1 := Conv(\mathcal{X}_0)$. We evaluate a given combination of prospects, $x^* \in \mathcal{X}_1$. Without loss of generality, we rank the scenarios in ascending order by the outcomes of the evaluated combination: $x_1^* \leq \cdots \leq x_S^*$.

Distinction is drawn between three closely related definitions of 'SD efficiency' which apply in this case. [Liz12b] and [KP15] provide further discussion of SD efficiency concepts.

Definition 3 (SD efficiency). An evaluated combination $x^* \in \mathcal{X}_1$ is efficient in terms of NSD, $N \geq 1$, if it is preferred to every alternative $x \in \mathcal{X}_1$ for some permissible utility function $u \in \mathcal{U}_N$:

$$\sum_{r=1}^{R} p_r u\left(x_{i,r}\right) \geq \sum_{r=1}^{R} p_r u\left(x_r\right) \ \forall x \in \mathcal{X}_1.$$

To implement this definition, [PK13] use the following equivalent definition, for $N \geq 2$:

Definition 3' (SD efficiency). An evaluated combination $x^* \in \mathcal{X}_1$ is efficient in terms of NSD, $N \geq 2$, if it obeys the Karush–Kuhn–Tucker first-order optimality conditions for some permissible utility function $u \in \mathcal{U}_N$:

$$\sum_{r=1}^{R} p_r u^1\left(x_r^*\right)\left(x_r^* - x_{j,r}\right) \geq 0,\ j = 1, \cdots, M. \tag{4}$$

This formulation was first introduced by [Post03] for SSD efficiency ($N = 2$) and extended by [PV07, Section IV] to TSD efficiency ($N = 3$). The formulation applies also for higher-degree efficiency criteria ($N \geq 4$). However, it would give a necessary but not sufficient condition for FSD efficiency ($N = 1$). [KP09] present an alternative formulation for FSD efficiency based on piece-wise constant utility functions which is equivalent to Definition 1 for $N = 1$.

The portfolio choice literature ([SY94], [DR03], [Kuos04], [RDM06], [AD15], [KP15], [Long16], [PK16]), generally uses a third definition of efficiency:

Definition 3" (SD efficiency). An evaluated combination $x^* \in \mathcal{X}_1$ is efficient in terms of NSD, $N \geq 1$, if it is not dominated by any alternative combination $x \in \mathcal{X}_1$, in terms of NSD.

For FSD efficiency ($N = 1$), Definition 3 is a sufficient but not necessary condition for Definition 3", as shown in [KP09]. Since both FSD efficiency definitions are already covered in detail in [Kuos04] and [KP09], our analysis focuses on NSD efficiency for $N \geq 2$. For this case, the above three efficiency definitions are equivalent, due to the saddle point property in the joint analysis of portfolio weights and risk preferences (see [Post03], Thm 1).

Consistent with the above equivalence relations, several existing portfolio optimization methods based on SSD ($N = 2$) search simultaneously over portfolio weights and utility functions ([KP15], [AD15], [Long16]). The future development of portfolio optimization methods based on NSD, $N \geq 3$, may benefit from the characterization of SD efficiency using piecewise polynomial functions in Section 6 below.

3 Local analysis

This section analyses utility functions and their derivatives on a given subinterval $[a, b] \subseteq \mathcal{D}$. The local analysis is relevant for the subintervals $[y_s, y_{s+1}]$, $s = 1, \cdots, S-1$, in Definition 2' and $[x_r, x_{r+1}]$, $r = 1, \cdots, R-1$, in Definition 3'. The next section analyses the global behavior on the entire outcomes domain $\mathcal{D}$.

For any given $u \in \mathcal{U}_N$, $N \geq 1$, consider the following decomposition of the nth derivative, $n = 0, \cdots, N-1$, based on Taylor expansions:

$$u^n(x) = \sum_{q=n}^{N-1} \frac{u^q(b)(x-b)^{q-n}}{(q-n)!} + \mathcal{R}_{u,n}(x;b); \tag{5}$$

$$\mathcal{R}_{u,n}(x;b) := -\frac{1}{(N-n-1)!} \int_x^b u^{(N)}(t)(x-t)^{N-n-1} dt. \tag{6}$$

The nth order derivative $u^{(n)}(x)$ consists of a Taylor polynomial and a remainder term $\mathcal{R}_{u,n}(x;b)$. The analytic challenge in this section is to characterize the relation between $\mathcal{R}_{u;n}(a;b)$ and $\mathcal{R}_{u;m}(a;b)$ for different orders ($n \neq m$).

Let $\boldsymbol{\nabla u}(x) := \left(u^0(x) \cdots u^{N-1}(x)\right)$. We introduce model variables $\boldsymbol{\alpha} :=$

$(\alpha_0 \cdots \alpha_{N-1})$ and $\boldsymbol{\beta} := (\beta_0 \cdots \beta_{N-1})$ to capture $\boldsymbol{\nabla u}(a)$ and $\boldsymbol{\nabla u}(b)$ for permissible utility functions. Consider the following linear combinations of the model variables:

$$\rho_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \frac{(N-n-1)!}{(a-b)^{N-n}} \left(\alpha_n - \sum_{q=n}^{N-1} \frac{\beta_q (a-b)^{q-n}}{(q-n)!} \right), \ n = 0, \cdots, N-1. \quad (7)$$

Importantly, $\rho_n(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is constructed to capture a normalization of the nth remainder term. Specifically, combining (5) and (7) yields

$$\rho_n(\boldsymbol{\nabla u}(a), \boldsymbol{\nabla u}(b)) = \frac{(N-n-1)!}{(a-b)^{N-n}} \mathcal{R}_{u,n}(a;b), \ n = 0, \cdots, N-1. \quad (8)$$

To capture the relation between the remainder terms of different orders, we will use the nth forward difference as defined through the binomial transform:

$$\sigma_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \sum_{k=0}^{n} (-1)^k \begin{pmatrix} n \\ k \end{pmatrix} \rho_k(\boldsymbol{\alpha}, \boldsymbol{\beta}), \ n = 0, \cdots, N-1. \quad (9)$$

The transform is self-inverse, so that the original levels can be regained from the differences in the following way:

$$\rho_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{k=0}^{n} (-1)^k \begin{pmatrix} n \\ k \end{pmatrix} \sigma_k(\boldsymbol{\alpha}, \boldsymbol{\beta}), \ n = 0, \cdots, N-1. \quad (10)$$

Consider the following joint restrictions for the model variables $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$:

$$(-1)^{n+1}\beta_n \geq 0,\ n = 0, \cdots, N-1; \tag{11}$$

$$(-1)^{N-n-1}\sigma_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) \geq 0,\ n = 0, \cdots, N-1; \tag{12}$$

$$(-1)^{N-n-1}\sum_{q=n}^{n+2m}\begin{pmatrix} 2m \\ q-n \end{pmatrix}\varphi^{q-n}\sigma_q(\boldsymbol{\alpha}, \boldsymbol{\beta}) \geq 0,$$

$$m = 1, \cdots, \left\lfloor \frac{N-1}{2} \right\rfloor; n = 0, \cdots, (N-2m-1); \forall\varphi \geq 0. \tag{13}$$

For $N = 1, 2$, inequalities (13) do not apply, as $\lfloor \frac{N-1}{2} \rfloor = 0$. In fact, (12) can be seen as the extension of (13) to $m = 0$. For $N = 3, 4$, we need to consider $m = 1$. For $N = 5, 6$, we need to consider also $m = 2$, and so forth for higher degrees ($N \geq 7$).

The above restrictions characterize the levels and derivatives of all permissible utility functions:

Lemma 1 (Local necessary conditions). For any given $u \in \mathcal{U}_N$, $N \geq 1$, we find that $\boldsymbol{\nabla u}(a)$ and $\boldsymbol{\nabla u}(b)$ obey (11), (12) and (13).

Lemma 2 (Local sufficient conditions). If $\boldsymbol{\alpha} = (\alpha_0 \cdots \alpha_{N-1})$ and $\boldsymbol{\beta} = (\beta_0 \cdots \beta_{N-1})$ obey (11), (12) and (13) for $N = 1, \cdots, 4$, then there exists $u \in \mathcal{U}_N^*$: $\boldsymbol{\nabla u}(a) = \boldsymbol{\alpha}$ and $\boldsymbol{\nabla u}(b) = \boldsymbol{\beta}$.

The proof of Lemma 2 in the Appendix is formulated in terms of functions $u \in \mathcal{U}_N^*$ that consist of $(N-1)$ polynomial pieces of degree $(N-1)$. An alternative proof (based on $\mathcal{U}_N$ rather than $\mathcal{U}_N^*$) appears as the proof of the sufficient condition of Theorem 1 in [Fang14].

Despite the complicated structure of the coefficients, constraints (12) and (13) are linear in the model variables $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The constraints in (13), which apply for $N \geq 3$, are however of infinite dimension due to the requirement $\forall\varphi \geq 0$.

For general $N \geq 3$, we can derive an approximate linear discretization by restricting the parameter φ to the unit interval:

Lemma 3 (Bounded parameter space). The conditions (13) are equivalent to

$$(-1)^{N-n-1} \sum_{q=n}^{n+2m} \binom{2m}{q-n} \varphi^{q-n} \sigma_q(\boldsymbol{\alpha}, \boldsymbol{\beta}) \geq 0; \tag{14}$$

$$(-1)^{N-n-1} \sum_{q=n}^{n+2m} \binom{2m}{q-n} \varphi^{n-q} \sigma_q(\boldsymbol{\alpha}, \boldsymbol{\beta}) \geq 0, \tag{15}$$

$$m = 1, \cdots, \left\lfloor \frac{N-1}{2} \right\rfloor; n = 0, \cdots, (N-2m-1); \forall \varphi \in [0,1].$$

(Without proof)

It is straightforward to develop a suitable discretization for the unit interval. Furthermore, for $N = 3, 4$ $(m = 1)$, an exact quadratic discretization exists:

Lemma 4 (Quadratic constraints). For $N = 3, 4$, the inequalities (13) are equivalent to

$$\sigma_{n+1}(\boldsymbol{\alpha}, \boldsymbol{\beta})^2 - \sigma_n(\boldsymbol{\alpha}, \boldsymbol{\beta})\sigma_{n+2}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \leq 0, \ n = 0, \cdots, N-3. \tag{16}$$

These quadratic constraints (16) are convex in the parameter space defined by inequalities (11), which require an alternating sign for the forward differences $\sigma_n(\boldsymbol{\alpha}, \boldsymbol{\beta})$, $n = 0, \cdots, N-1$. Hence, the finite quadratic system $\{(11), (12), (16)\}$ is convex.

4 Global analysis

To implement SD optimality and SD efficiency, we will now consider the case with T outcomes z_t, $t = 1, \cdots, T$, $z_1 \leq \cdots \leq z_T$. The outcomes partition the interval $[z_1, z_T]$ into sub-intervals $[a_t, b_t] := [z_t, z_{t+1}]$, $t = 1, \cdots, T-1$, where $a_{t+1} = b_t = z_{t+1}$, $t = 1, \cdots, T-1$.

Proposition 1 (Global necessary conditions). For any utility function $u \in \mathcal{U}_N$, $N \geq 1$, and outcomes $z_1 \leq \cdots \leq z_T$, $(\boldsymbol{\nabla u}(z_t), \boldsymbol{\nabla u}(z_{t+1}))$, $t = 1, \cdots, T-1$, obey inequalities (11), (12) and (13).

For every sub-interval, we use parameter vectors to capture $\boldsymbol{\nabla u}(a_t)$ and $\boldsymbol{\nabla u}(b_t)$. Since the sub-intervals are connected, or $a_{t+1} = b_t = z_{t+1}$, $t = 1, \cdots, T-1$, we find $\boldsymbol{\nabla u}(a_t) = \boldsymbol{\nabla u}(b_{t-1})$, $t = 2, \cdots, T-1$. We therefore only need a single set of parameters $\boldsymbol{\alpha}_t$, $t = 1, \cdots, T$, where $\boldsymbol{\alpha}_t = \boldsymbol{\nabla u}(a_t)$ $t = 1, \cdots, T-1$, and $\boldsymbol{\alpha}_T = \boldsymbol{\nabla u}(b_{T-1})$.

Proposition 2 (Global sufficient conditions). For given outcomes $z_1 \leq \cdots \leq z_T$ and degree $N = 1, 2, 3, 4$, if a given set of parameters $\boldsymbol{\alpha}_t = (\alpha_{t,0} \cdots \alpha_{t,N-1})$, $t = 1, \cdots, T$, satisfy the inequalities (11), (12) and (13) for every $(\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_{t+1})$, $t = 1, \cdots, T-1$, then there exists $u \in \mathcal{U}_N^*$: $\boldsymbol{\nabla u}(z_t) = \boldsymbol{\alpha}_t$, ; $t = 1, \cdots, T$.

Using $(N-1)$ polynomial pieces for every subinterval, we find that $\mathcal{U}_N^*$ can be represented by $(N-1)(T-1)$ polynomial pieces of degree $(N-1)$.

5 SD Optimality conditions

The inequalities (11), (12) and (13) are linear in the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Hence, the inequalities (2) are also linear in these parameters. We use $T = S$ and $z_t = y_s$. Applying Proposition 1 and Proposition 2 to the SD optimality conditions (2), we find a linear system for SD optimality:

Theorem 1 (SD optimality). An evaluated prospect x_i, $i = 1, \cdots, M$, is optimal in terms of NSD, $N \geq 1$, only if there exists a non-zero solution for the following system of inequalities:

$$\sum_{s=1}^{S} \alpha_{0,s} \left(p_{i,s}^* - p_{j,s}^*\right) \geq 0,\ j = 1, \cdots, M; \tag{17}$$

$$(-1)^{n+1} \alpha_{n,s} \geq 0,\ n = 0, \cdots, N-1; s = 1, \cdots, S;$$

$$(-1)^{N-n-1} \sigma_n \left(\boldsymbol{\alpha}_s, \boldsymbol{\alpha}_{s+1}\right) \geq 0,\ n = 0, \cdots, N-1; s = 1, \cdots, S-1;$$

$$(-1)^{N-n-1} \sum_{q=n}^{n+2m} \begin{pmatrix} 2m \\ q-n \end{pmatrix} \varphi^{q-n} \sigma_q \left(\boldsymbol{\alpha}_s, \boldsymbol{\alpha}_{s+1}\right) \geq 0,$$

$$m = 1, \cdots, \left\lfloor \frac{N-1}{2} \right\rfloor; n = 0, \cdots, (N-2m-1); s = 1, \cdots, S-1; \forall \varphi \geq 0.$$

For $N = 1, 2, 3, 4$, these inequalities are also sufficient conditions.
(Without proof)

We must exclude zero solutions, or $\alpha_{n,s} = 0$ for all $n = 1, \cdots, N-1$ and $s = 1, \cdots, S$, to avoid the trivial utility function $u(x) = c$, $\forall x \in \mathcal{D}$, or an indifferent decision maker.

For $N = 1, 2$, the system consists of a finite number of linear inequalities. For $N \geq 3$, we can obtain an approximate linear discretization using Lemma 3 and a discretization of the unit interval for the parameter φ.

We can specify optimization problems to test the linear systems. Details such as the orientation of the objective function and the normalization of the variables depend on the application at hand. Appendix B discusses the CQP problem for the optimality test that we use in the numerical example in Section 7 below.

In addition, for $N = 3, 4$, we may use Lemma 4 to find a finite number of quadratic inequalities for TSD and FOSD optimality. These inequalities can be tested using CQP.

6 SD efficiency conditions

In a similar way, we can derive linear or quadratic systems for SD efficiency. We could apply Proposition 1 and Proposition 2 directly to the utility function $u(x)$ in Definition 3'. However, this approach would introduce redundancies, as the efficiency conditions (3) do not require the utility levels $u(a)$ and $u(b)$.

A more computationally efficient approach applies our results to the negative of the marginal utility function, $n(x) := -u^1(x)$. Since $n(x) \in \mathcal{U}_{N-1}$, this approach allows us implement Nth degree SD efficiency in the same way as $(N-1)$th degree SD optimality.

To remove the utility levels from the analysis, we modify the definition of the forward differences as follows:

$$\varsigma_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \sum_{k=1}^{n} (-1)^k \begin{pmatrix} n \\ k \end{pmatrix} \rho_k(\boldsymbol{\alpha}, \boldsymbol{\beta}), \ n = 1, \cdots, N-1. \tag{18}$$

Furthermore, we use $T = R$ and $z_t = x_r$.

We can now present the analogue of Theorem 1 for SD efficiency:

Theorem 2 (SD efficiency). An evaluated combination $x^* \in \mathcal{X}_1$ is efficient in terms of NSD, $N \geq 2$, only if there exists a non-zero solution for the following system of inequalities:

$$\begin{gathered}
\sum_{r=1}^{R} p_r \alpha_{1,r} \left(x_r^* - x_{j,r}\right) \geq 0, \; j = 1, \cdots, M; \qquad (19) \\
(-1)^{n+1} \alpha_{n,r} \geq 0, \; n = 1, \cdots, N-1; r = 1, \cdots, R; \\
(-1)^{N-n} \varsigma_n \left(\boldsymbol{\alpha}_r, \boldsymbol{\alpha}_{r+1}\right) \geq 0, \; n = 1, \cdots, N-1; r = 1, \cdots, R-1; \\
(-1)^{N-n} \sum_{q=n}^{n+2m} \left(\begin{array}{c} 2m \\ q-n+1 \end{array} \right) \varphi^{q-n} \varsigma_q \left(\boldsymbol{\alpha}_r, \boldsymbol{\alpha}_{r+1}\right) \geq 0, \\
m = 1, \cdots, \left\lfloor \frac{N-2}{2} \right\rfloor ; n = 1, \cdots, (N-2m-1) \, ; r = 1, \cdots, R-1; \forall \varphi \geq 0.
\end{gathered}$$

For $N = 2, 3, 4, 5$, these inequalities are also sufficient conditions.
(Without proof)

For SSD and TSD ($N = 2, 3$), the system is finite and linear. For $N \geq 4$, we may obtain a linear approximate discretization along the lines of Lemma 3 or, for $N = 4, 5$, a convex quadratic exact discretization along the lines of Lemma 4.

7 Illustration ($N = 3$)

This section illustrates our analysis for the important case of $N = 3$. Figure 1 shows bounds on $\boldsymbol{\nabla u}(x)$, $u \in \mathcal{U}_3$, obtained for particular values of the model variables $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Let $[a, b] = [0.8, 1.2]$ and $k(x) = -\exp(-4x)$.

Without specifying $\boldsymbol{\alpha}$, the only restrictions on $\boldsymbol{\beta}$ are the alternating signs. In our example, we set $\boldsymbol{\beta} = \boldsymbol{\nabla k}(1.2)$. For α_2 and β_2 to represent the curvature $u^2(a)$ and $u^2(b)$, $u \in \mathcal{U}_3$, we must have that $\alpha_2 \leq \beta_2$. Fixing the values for α_n, $n = 2, 1, 0$, introduces additional restrictions. Suppose that we select a feasible value for α_2, say $\alpha_2 = k^2(a) = -16\exp(-3.2)$. For α_1 and β_1 to represent the slope $u^1(a)$ and $u^1(b)$ of some $u \in \mathcal{U}_3$, the mean-value theorem implies

$$\alpha_1 \leq \beta_1 + \alpha_2(a - b); \qquad (20)$$

$$\alpha_1 \geq \beta_1 + \beta_2(a - b). \qquad (21)$$

Next, we select a specific feasible value for α_1, say $\alpha_1 = k^1(a) = 4\exp(-3.2)$. This choice further narrows the range of relevant functions. The remaining functions can be characterized by a lower envelope $g(x)$ and an upper envelope $h(x)$, which are formally defined in (37) and (39) in the Appendix. These two extreme functions and their derivatives are shown in Figure 1 as the dashed lines (lower envelope) and the dotted lines (upper envelope). In our specific example, $\gamma \approx -0.33$ and $\theta \approx 0.95$.

For α_0 and β_0 to represent the levels $u(a)$ and $u(b)$, $u \in \mathcal{U}_3$, we must have

$$\alpha_0 \geq g(a) = \beta_0 + \beta_1(a-b) + \frac{1}{2}\left(\alpha_1 - \beta_1\right)(a-b); \tag{22}$$

$$\alpha_0 \leq h(a) = \beta_0 + \beta_1(a-b) + \frac{1}{2}\beta_2(a-b)^2 + \frac{1}{2}(\alpha_2 - \beta_2)(a-\theta)^2. \tag{23}$$

It is easy to verify that inequalities (20), (21) and (22) amount to the three conditions in (12) and the inequality (23) amounts to (16) for $N = 3$.

The above analysis illustrates the necessary condition (Lemma 1). We can also use the example to illustrate sufficiency (Lemma 2). Suppose that we select a feasible value for α_0, say $\alpha_0 = k(a) = -\exp(-3.2)$. We can find $u \in U_3^*$: $(\boldsymbol{\nabla u}(a), \boldsymbol{\nabla u}(b)) = (\boldsymbol{\alpha}, \boldsymbol{\beta})$, by taking the mixture $f(x) = wg(x) + (1-w)h(x)$, $w \in [0,1]$, that gives $f(a) = \alpha_0$. By construction, the mixed function is permissible ($f \in \mathcal{U}_3^*$) and $\boldsymbol{\nabla f}(a) = \boldsymbol{\alpha}$ and $\boldsymbol{\nabla f}(b) = \boldsymbol{\beta}$). In Figure 1, the resulting function and its derivatives are shown as the solid lines in the three panels.

Figure 2 continues the example to illustrate that local monotonicity ensures global monotonicity (Proposition 2). The graph combines the local results for three subintervals: $[a_1, b_1] = [0.4, 0.8]$, $[a_2, b_2] = [0.8, 1.2]$ and $[a_3, b_3] = [1.2, 1.6]$. In this case, the combined function consist of $(N-1)(T-1) = 6$ quadratic pieces.

By contrast, the [PK13] approach would consider only utility functions with $(T-1) = 3$ quadratic pieces, which can lead to false rejections of optimality or efficiency. To illustrate the distinction between the two approaches, consider the following example with two prospects ($M = 2$) and four possible outcomes ($T = 4$):

t	y_t	$p^*_{1,t}$	$p^*_{2,t}$
1	0.4	0.32	0.34
2	0.8	0.19	0.12
3	1.2	0.31	0.42
4	1.6	0.18	0.12

In this case, x_1 does not dominate x_2 by TSD, as may be verified using a simple pairwise TSD test. Given that non-dominance is equivalent to optimality for $M = 2$, we conclude that x_2 is TSD optimal. We may illustrate the distinction between the [PK13] approach and our approach by testing whether x_2 is TSD optimal using both approaches.

The [PK13] approach falsely classifies x_2 as non-optimal, because x_1 achieves a higher expected utility than x_2 for all three-piece quadratic functions with kinks at the subinterval boundary points $x = 0.8, 1.2$. The optimal solution to the relevant LP problem ([PK13], Eq. 15) is given by

$$(u(0.4), u(0.8), u(1.2), u(1.6)) = (-1.08, -0.48, -0.12, 0.00); \tag{24}$$

$$(u^1(0.4), u^1(0.8), u^1(1.2), u^1(1.6)) = (1.80, 1.20, 0.60, 0.00). \tag{25}$$

This solution corresponds to the following one-piece quadratic utility function:

$$u(x) = -1.905 + 2.381x - 0.744x^2. \tag{26}$$

Even for this 'most favorable' function, x_1 achieves a higher expected utility than x_2. Unfortunately, the [PK13] approach does not consider the possibility that the utility function kinks in the interior of the subintervals.

We may implement our approach using the CQP problem that is described in Appendix B. The optimal value of the objective function is zero, which means that x_2 is classified as TSD optimal. The optimal solution is not unique. An example of an optimal solution is given by

$$(u(0.4), u(0.8), u(1.2), u(1.6)) = (-1.01, -0.36, -0.04, 0.00); \tag{27}$$

$$(u^1(0.4), u^1(0.8), u^1(1.2), u^1(1.6)) = (2.02, 1.21, 0.40, 0.00). \tag{28}$$

This solution corresponds to the following two-piece quadratic utility function with a kink at the interior point $x = 1.4$:

$$u(x) = \begin{cases} -1.01\,(1.4 - x)^2 & x \leq 1.4 \\ 0 & x > 1.4 \end{cases}. \tag{29}$$

For this function, x_2 achieves a higher expected utility than x_1. Thus, our approach correctly classifies x_2 as TSD optimal, based on a permissible function which is ignored by the [PK13] approach.

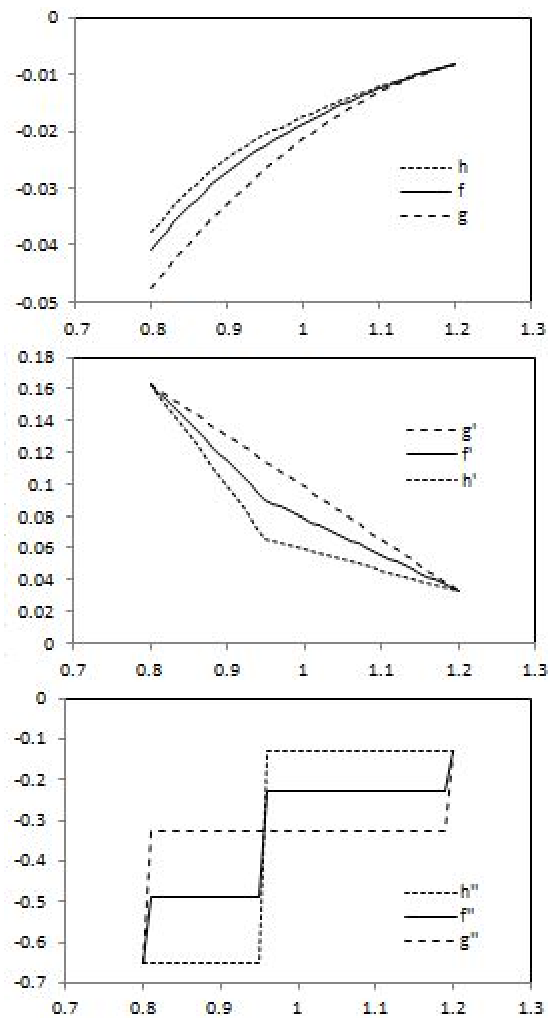

Figure 1 Local analysis for TSD

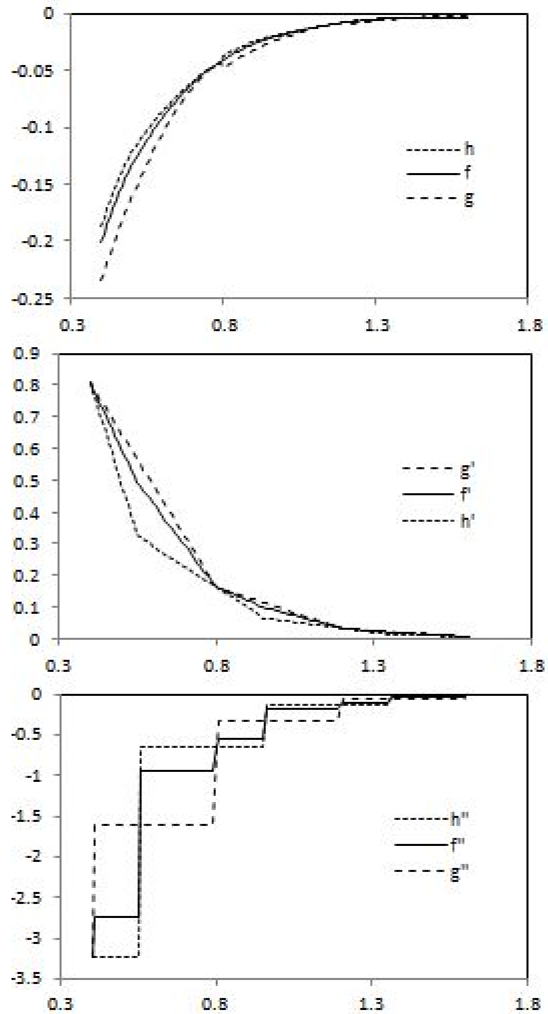

Figure 2 Global analysis for TSD

8 Discussion

Our analysis allows us to further analyze the approximation by [PK13]. Their Theorem 1 implicitly assumes that $\mathcal{R}_{u,n}(x;b) = c\frac{1}{(N-n-1)!}(x-b)^{N-n-1}$, $n = 0, \cdots, N-2$. By the Lagrange form of the remainder term, this representation is correct for a given n if we set $c = u^{N-1}(\xi_n)$, where $\xi_n \in [a,b]$.

Unfortunately, the points ξ_n, $n = 0, \cdots, N-2$, are generally not identical unless the utility function consists of one $(N-1)$th degree polynomial piece on the relevant subinterval $[a, b]$.

The approximation is perfect for FSD and SSD optimality tests and SSD and TSD efficiency tests. These tests can be formulated in terms of piece-wise constant or piece-wise linear utility functions or marginal utility functions. However, approximation error can arise for SD optimality tests of degree $N \geq 3$ and SD efficiency tests of degree $N \geq 4$.

The approximation error disappears as $(b-a) \to 0$. Therefore, the flaw has no material consequences if the outcomes z_t, $t = 1, \cdots, T$, represent a fine partition of the outcomes domain. The application of [PK13] uses large financial data sets with a dense empirical distribution. In this situation, our revision has no material effect. Nevertheless, our revision can lead to improvements for higher-degree tests if the partition of the outcomes domain is more coarse, for example, in behavioral choice experiments.

Our revision introduces model variables for the values of all relevant derivatives at all relevant outcome levels. Despite the additional variables and constraints, the problem dimensions of (17) and (19) remain linear in the number of scenarios. The problems remain relatively small for typical applications, even for a fine discretization of the unit interval for the parameter φ. Also a CQP formulation based on Lemma 4 is inexpensive with modern-day computer hardware and solver software.

Our strongest results are obtained for SD optimality based on $N = 1, 2, 3, 4$ and SD efficiency based on $N = 2, 3, 4, 5$. For SD optimality and efficiency of higher degrees, we produce only necessary conditions. We did not pursue stronger results, because degree $N \geq 5$ seems to have limited practical use. Restrictions on the signs of the higher-order derivatives generally have minimal effects on the flexibility to model the utility levels (in Definition 2') and marginal utility levels (in Definition 3').

By contrast, restrictions on the Pratt-Arrow coefficient of absolute risk aversion (DARA SD and SDWRF), level of the lower-order derivatives (ASD) and Kimball's coefficient of absolute prudence (StSD) tend to be more powerful than restrictions on the signs of the higher-order derivatives ([BP97], [PK13], [PFK15], [PP16]). Our results for $N = 1, 2, 3, 4$ are relevant in this context, as DARA SD and StSD maintain the standard TSD and FOSD restrictions and, in addition, all these relations (DARA SD, SDWRF, ASD, StSD) require the values of the derivatives of various orders.

Our approach also allows us to implement the [MSTW16] SD rule, which falls between FSD and SSD. Using $0 \leq C \leq 1$ for an anti-index of greediness, [MSTW16] introduce the following notion of generalized concavity:

$$u^1(z_1) \geq Cu^1(z_2) \geq 0, \ \forall z_1, z_2 : z_1 \leq z_2. \tag{30}$$

In the special case of $C = 0$, this condition amounts to non-satiation (FSD); for $C = 1$, we obtain risk aversion (SSD). The case with $0 < C < 1$ falls in between of these two special cases.

We know that

$$\sigma_0(\boldsymbol{\alpha}_s, \boldsymbol{\alpha}_{s+1}) = \frac{u(x_s) - u(x_{s+1})}{x_s - x_{s+1}} = u^1(z), \tag{31}$$

for some $z \in [x_s, x_{s+1}]$. It is therefore possible to extend our linear system for FSD optimality ((17); $N = 1$) to the following system for the new SD rule:

$$\begin{gathered} \sum_{s=1}^{S} \alpha_{0,s} \left(p_{i,s}^* - p_{j,s}^*\right) \geq 0, \ j = 1, \cdots, M; \\ \alpha_{0,s} \leq 0, \ s = 1, \cdots, S; \\ \alpha_{1,s} \geq C\sigma_0(\boldsymbol{\alpha}_s, \boldsymbol{\alpha}_{s+1}) \geq 0, \ s = 1, \cdots, S-1; \\ \sigma_0(\boldsymbol{\alpha}_s, \boldsymbol{\alpha}_{s+1}) \geq C\alpha_{1,s+1} \geq 0, \ s = 1, \cdots, S-1. \end{gathered} \tag{32}$$

Our analysis of SD efficiency can however not be extended in this way, because Definition 3' is based on the Karush–Kuhn–Tucker conditions, which generally are not sufficient if we allow for risk seeking.

9 Empirical application

We can use our linear system (19) to analyze market portfolio efficiency along the lines of [PK13]. We compare the CRSP all-share index with actively managed stock portfolios that are formed, and periodically rebalanced, based on publicly available stock-level information. The analysis also includes a riskless asset with return equal to the time-series average of the T-bill yield in the relevant sample period.

We consider six different sets of portfolios from the data library of Kenneth

French: (i) ten portfolios formed on market capitalization of equity (ME); (ii) 30 portfolios formed on four-digit Standard Industrial Classification (SIC) codes; (iii) 25 portfolios formed on ME and book-to-market equity ratio (BM); (iv) 25 portfolios formed on ME and the return in the past month (R1-1); (v) 25 portfolios formed on ME and the return in the eleven months before the past month (R2-12); (vi) 25 portfolios formed on ME and the return in the four years before the past year (R13-60).

We analyze gross value-weighted portfolio returns for holding periods of one month ($H = 1$), one quarter ($H = 3$) and one year ($H = 12$) from the first available observation, depending on the data set, in the late 1920s or early 1930s, to the end of 2015.

We design optimization problems for system (19) for NSD efficiency by degree $N = 2, 3, 4, 5$. Appendix B specifies the relevant LP and CQP problems. In addition to the NSD efficiency tests, we apply efficiency tests based on DSD, SDWRF, ASSD and StSD, using additional model variables and constraints from the existing literature. We also apply a mean-variance (MV) efficiency test based on a linear and decreasing marginal utility function.

The time-series observations are interpreted as scenarios with equal probabilities ($p_r = R^{-1}$). In this application, $u^1\left(x_r^*\right), r = 1, \cdots, R$, can be interpreted as a stochastic discount factor (SDF) that equals the marginal utility of wealth for a representative investor. All tests are normalized such that the sample mean of the SDF equals unity: $R^{-1} \sum_{r=1}^{R} u^1\left(x_r^*\right) = 1$.

Since all base assets have a strictly positive weight in the index, the first-order optimality condition (3) must hold with equality and the violations $\varepsilon_j := \sum_{r=1}^{R} p_r u^1\left(x_r^*\right)\left(x_{j,r} - x_r^*\right)$, $j = 1, \cdots, M$, can be interpreted as pricing errors. The objective of our optimization problem is to minimize (across all permissible utility functions) the maximum (across the base assets) of the pricing errors. We prefer this mini-max criterion because it allows for a straightforward economic interpretation of the objective as the largest abnormal return that can be achieved without leverage or short selling.

For statistical inference, we use a bootstrap procedure that repeatedly applies the efficiency test to random pseudo-samples. Under the assumption of identical and independently distributed (i.i.d.) time-series returns, the empirical return distribution is a consistent estimator of the population distribution, and bootstrap samples can simply be obtained by randomly sampling with replacement from the empirical distribution.

In order to obtain consistent p-values, it is important to re-center the bootstrap process so that it obeys the null hypothesis ([HH96]). Our null hypothesis is portfolio efficiency and the empirical violations of this null hypothesis are the estimated pricing errors $\hat{\varepsilon}_j := R^{-1}\sum_{r=1}^{R}\alpha_{1,r}\left(x_{j,r}-x_r^*\right), r=1,\cdots,R$; $j=1,\cdots,M$. We can therefore re-center the bootstrap process by subtracting the estimated pricing errors from every observation: $\hat{x}_{j,r} := x_{j,r}-\hat{\varepsilon}_j$, $r=1,\cdots,R$; $j=1,\cdots,M$.

We implement the bootstrap by generating pseudo-samples of the same size as the original sample through random draws with replacement from the recentered version of the original sample, and test efficiency in every pseudo-sample. Finally, we compute critical values for the original test statistic from the percentiles of the bootstrap distribution.

In a specialized study of bootstrap inference on SD efficiency, [ST2010] recommend 300 pseudo-samples as a reasonable compromise between accuracy, time and computer constraints for a similar application. To be on the safe side, we used 10,000 pseudo-samples, at the cost of additional computer time.

Table I summarizes the estimation results. Perhaps surprisingly, we cannot reject MV efficiency of the market portfolio in most of the data sets. The relatively large differences in average return imply a high estimated market risk premium. The MV SDF therefore takes negative values for the largest market upswings, which violates the no-arbitrage principle (a problem discussed by [DI82]) and lowers the estimated pricing errors of high-beta investment portfolio. The MV SDF also does not penalize the negative skewness of the market index.

The SSD efficiency test yields even smaller pricing errors in most data sets. The median test statistic falls by 67 basis points (bps) compared with the MV test. However, the SSD SDF is a far cry from a well-behaved marginal utility function. The SDF tends to show large discontinuous jumps and concave segments. This pattern is not consistent with decreasing risk aversion and casts doubt on the economic meaning of the SSD results.

The TSD criterion ($N=3$) imposes prudence (skewness preference) and avoids non-convexity of the SDF. The median test statistic increases by 99 bps compared with SSD. The TSD criterion also tends to be stronger than the MV criterion because it excludes arbitrage opportunities and accounts for systematic skewness.

The incremental effect of restricting the higher-order derivatives diminishes. The FOSD criterion ($N=4$) assumes temperance (kurtosis aversion) and increases the median test statistic by 14 bps. The incremental effect of imposing edginess (FISD; $N=5$) is limited to just a few bps. The test results for NSD of even higher degree ($N \geq 6$) are not distinguishable from the FISD results.

Although the higher-degree SDFs ($N \geq 3$) are convex, they exhibit large linear segments or local risk neutrality. The DSD criterion imposes DARA (or log-convexity of the SDF). This assumption has more discriminating power than sign restrictions for the higher-order derivatives, witness an 71 bps increase of the median test statistic compared with TSD. StSD imposes DAP in addition to DARA. The incremental effect of this assumption is just a few bps in the median sample.

SDWRF and ASSD lead to large increases in the test statistic compared with SSD but the results are incomparable with TSD and DSD; the ranking of these decision rules depends on the specific data set. This is not entirely surprising, because restricting the level of risk aversion or ARA differs fundamentally from restricting the direction of risk aversion or ARA.

As a robustness test, we repeated our analysis after excluding the first size quintile (in the five data sets that sort stocks on ME) and the early sub-period before 1963, common robustness tests in the empirical asset pricing literature. The full-sample results and conclusions are robust to these exclusions. The median value of the test statistic decreases by tens of bps, but the effect of restricting higher-order risk aversion is comparable with that in the full sample.

A robust conclusion seems that SSD is too weak for meaningful investment analysis. By contrast, higher-order SD rules, as well as modifications of SSD based on SDWRF and ASSD, seem useful as a complement to MV analysis. In particular restrictions on the level or direction of risk aversion and the ARA coefficient are effective to increase the discriminating power of the analysis. It adds to the mounting evidence against market portfolio efficiency that the estimated pricing errors based on higher-order SD rules are even larger and more significant than standard MV estimates.

Table I Market portfolio efficiency

Shown are results for testing efficiency of the passive market index relative to various sets of active stock portfolios and the one-month T-bill for holding periods of $H = 1, 3, 12$ months. For interpretation and comparability, the pricing errors for an holding period of H months are 'annualized' by multiplication with $(12/H)$. We considered the following decision criteria: Mean-Variance (MV) dominance; Second-degree Stochastic Dominance (SSD); Third-degree Stochastic Dominance (TSD); Fourth-degree Stochastic Dominance (FOSD); Fifth-degree Stochastic Dominance (FISD); Decreasing Absolute Risk Aversion (DARA) SD; Standard Stochastic Dominance (StSD); Almost Second-degree Stochastic Dominance (ASD); Stochastic Dominance With respect to a Function (SDWRF). We analyze gross value-weighted portfolio returns from the first available date in Kenneth French data library until the end of 2015. Bootstrap critical values are based on 10,000 pseudo-samples from the original sample after re-centering the means. We use *, ** and *** to denote a statistical confidence level of 90%, 95% and 99%, respectively.

Data set	H	MV	SSD	TSD	FOSD	FISD	DSD	StSD	ASSD	SDWRF
	1	0.69	0.65	1.69	1.72	1.72	2.03	2.06	0.89	2.01
ME10	3	0.67	0.62	1.26	1.37	1.38	1.67	1.70	0.85	1.17
	12	1.76	0.78	1.811	1.87	1.87	2.18	2.21	0.97	2.00
	1	3.48	3.42	3.46	3.47	3.47	3.46	3.47	3.46	3.44
Ind30	3	3.60	3.19	3.60	3.60	3.60	3.64	3.65	3.24	3.29
	12	4.13	3.74	4.13	4.13	4.13	4.15	4.15	3.87	3.85
	1	4.00**	2.70*	4.10***	4.20***	4.21***	4.38***	4.44***	3.36**	4.91*
5MEx5BM	3	2.44	2.23	3.72**	3.86**	3.89**	4.09**	4.12**	3.04*	3.81*
	12	3.77	3.03*	4.02**	4.10**	4.10**	5.13**	5.14**	3.45*	4.27*
	1	8.00***	6.80***	8.33***	8.74***	8.74***	9.09***	9.18***	7.64***	12.01***
5MEx5R1-1	3	4.23	5.06***	8.04***	8.60***	8.81***	9.04***	9.15***	6.78***	10.68***
	12	6.54*	7.52**	10.48***	11.41***	11.63***	17.26***	17.27***	9.65**	16.15***
	1	7.38***	6.17***	7.25***	7.37***	7.37***	7.25***	7.39***	6.70***	8.57***
5MEx5R2-12	3	6.72	5.41***	6.88***	7.14***	7.19***	7.12***	7.24***	5.97***	7.82***
	12	6.31***	4.88***	6.92***	7.46***	7.46***	9.54***	9.56***	5.66***	8.82***
	1	4.72**	3.38**	4.88**	5.05***	5.06***	5.28***	5.33***	3.99**	5.98**
5MEx5R13-60	3	3.23	3.07*	4.66**	4.86**	4.92**	5.08**	5.15**	4.01**	4.95**
	12	6.66	5.54**	7.08**	7.33**	7.42**	8.34**	8.42***	6.09**	7.63**
Median		4.07	3.40	4.39	4.53	4.57	5.10	5.14	3.93	4.93

Appendix A: Formal proofs

Proof of Lemma 1. (11) follows directly from the definition of $\mathcal{U}_N$ in (1). Proving (12) and (13) involves some calculus. Using (9), (8) and (6),

$$\begin{aligned}\sigma_n\left(\boldsymbol{\nabla u}(a), \boldsymbol{\nabla u}(b)\right) &= \sum_{k=0}^{n}(-1)^k \binom{n}{k} \rho_k\left(\boldsymbol{\nabla u}(a), \boldsymbol{\nabla u}(b)\right)\\ &= \sum_{k=0}^{n}(-1)^k \binom{n}{k} \left(\frac{(N-k-1)!}{(a-b)^{N-k}}\mathcal{R}_{u,k}(a;b)\right)\\ &= \sum_{k=0}^{n}(-1)^k \binom{n}{k} \left(\frac{\int_a^b -u^N(t)(a-t)^{N-k-1}dt}{(a-b)^{N-k}}\right)\\ &= \frac{\int_a^b -u^N(t)(a-t)^{N-n-1}\sum_{k=0}^{n}(-1)^k \binom{n}{k}(a-t)^{n-k}(a-b)^k dt}{(a-b)^N}\\ &= \frac{\int_a^b -u^N(t)(a-t)^{N-n-1}(a-t-a+b)^n dt}{(a-b)^N}\\ &= \int_a^b -u^N(t)\frac{(a-t)^{N-n-1}(b-t)^n}{(a-b)^N}dt. \end{aligned} \tag{33}$$

Condition (12) now follows from

$$\operatorname{sgn}\left(-u^N(t)\frac{(a-t)^{N-n-1}}{(a-b)^N}\right) = (-1)^{(N-n-1)}. \tag{34}$$

Entering (33) in the constraint (13), and rearranging terms, yields

$$\begin{aligned}&(-1)^{(N-n-1)}\sum_{q=n}^{n+2m}\binom{2m}{q-n}\varphi^{q-n}\sigma_q\left(\boldsymbol{\nabla u}(a), \boldsymbol{\nabla u}(b)\right) =\\ &(-1)^{(N-n-1)}\sum_{q=n}^{n+2m}\binom{2m}{q-n}\varphi^{q-n}\left(\int_a^b -u^N(t)\frac{(a-t)^{N-n-1}(b-t)^n}{(a-b)^N}dt\right) =\\ &(-1)^{(N-n-1)}\int_a^b -u^N(t)\frac{(a-t)^{N-n-1-2m}(b-t)^n\left((a-t)^n+\varphi(b-t)\right)^{2m}dt}{(a-b)^N} \geq 0.\end{aligned} \tag{35}$$

The RHS is non-negative, because

$$\operatorname{sgn}\left(-u^N(t)\frac{(a-t)^{N-n-1-2m}}{(a-b)^N}\right) = (-1)^{(N-n-1)}.\square \tag{36}$$

Proof of Lemma 2. We will prove the lemma using piecewise polynomial functions $u \in \mathcal{U}_N^*$. For $N = 1, 2$, the results are straightforward based on the piecewise constant and piecewise linear functions of [KP09] and [Post03].

For $N = 3$, consider the following two piecewise quadratic functions:

$$g(x) := \beta_0 + \beta_1(x-b) + \frac{1}{2}\gamma(x-b)^2; \tag{37}$$

$$\gamma := \left(\frac{\alpha_1 - \beta_1}{a-b}\right); \tag{38}$$

$$h(x) := \begin{cases} \beta_0 + \beta_1(x-b) + \frac{1}{2}\beta_2(x-b)^2 + \frac{1}{2}(\alpha_2 - \beta_2)(x-\theta)^2 & x \leq \theta \\ \beta_0 + \beta_1(x-b) + \frac{1}{2}\beta_2(x-b)^2 & x > \theta \end{cases}; \tag{39}$$

$$\theta := a + \frac{\alpha_1 - \beta_1 - \beta_2(a-b)}{\beta_2 - \alpha_2}. \tag{40}$$

Restrictions (12, $n = 1$) and (16) can be reformulated as follows

$$\alpha_0 \geq \beta_0 + \beta_1(a-b) + \frac{1}{2}\left(\alpha_1 - \beta_1\right)(a-b) = g(a); \tag{41}$$

$$\alpha_0 \leq \beta_0 + \beta_1(a-b) + \frac{1}{2}\beta_2(a-b)^2 + \frac{1}{2}(\alpha_2 - \beta_2)(a-\theta)^2 = h(a). \tag{42}$$

Since $h(a) \geq \alpha_0 \geq g(a)$, we can find a mixture $f(x) = wg(x) + (1-w)h(x)$, $w \in [0, 1]$, that gives $f(a) = \alpha_0$. The mixture function is two-piece quadratic and permissible: $g, h \in \mathcal{U}_3^* \Rightarrow f \in \mathcal{U}_3^*$. The parameters γ and θ are set such that $g^n(z) = h^n(z) = u^n(z)$, for $z = a, b$ and $n = 1, 2$. Hence, we find $(\boldsymbol{\nabla f}(a), \boldsymbol{\nabla f}(b)) = (\boldsymbol{\alpha}, \boldsymbol{\beta})$.

For $N = 4$, we can apply the same reasoning using two-piece cubic functions, but the notation becomes more cumbersome. This case can be characterized in a compact way by the following two-piece step functions for the *third-order* derivative:

$$g^3(x) := \begin{cases} \alpha_3 & a \le x < \theta_1 \\ \gamma_1 & \theta_1 \le x < b \\ \beta_3 & x = b \end{cases}; \tag{43}$$

$$\gamma_1 := \alpha_3 - \frac{1}{2}\frac{(\alpha_2 - \beta_2 - \alpha_3(a-b))^2}{\alpha_1 - \beta_1 - \alpha_2(a-b) + \frac{1}{2}\alpha_3(a-b)^2}; \tag{44}$$

$$\theta_1 := a - 2\frac{\alpha_1 - \beta_1 - \frac{1}{2}(\alpha_2 + \beta_2)(a-b)}{\alpha_2 - \beta_2 - \alpha_3(a-b)}. \tag{45}$$

$$h^3(x) := \begin{cases} \alpha_3 & x = a \\ \gamma_2 & a < x \le \theta_2 \\ \beta_3 & \theta_2 < x \le b \end{cases}; \tag{46}$$

$$\gamma_2 := \frac{1}{2}\frac{(\alpha_2 - \beta_2 - \beta_3(a-b))^2}{\alpha_1 - \beta_1 - \beta_2(a-b) - \frac{1}{2}\beta_3(a-b)^2} + \beta_3; \tag{47}$$

$$\theta_2 := a - 2\frac{\alpha_1 - \beta_1 - \beta_2(a-b) - \frac{1}{2}\beta_3(a-b)^2}{\alpha_2 - \beta_2 - \beta_3(a-b)}. \tag{48}$$

The parameters γ_1, γ_2, θ_1 and θ_2 are set such that $g^n(z) = h^n(z) = u^n(z)$, for $z = a, b$ and $n = 1, 2, 3$. Integrating these step functions thrice gives the two relevant extremes for the utility functions. Mixing the two extremes yields a three-piece cubic utility function $f \in \mathcal{U}_4^*$ that takes the correct values for the levels and derivatives.□

Proof of Lemma 4. For $N = 3, 4$, we have that $\left\lfloor \frac{N-1}{2} \right\rfloor = 1$, and we need to consider only $m = 1$. For $m = 1$, the inequalities (13) are quadratic expressions of the parameter φ:

$$0 \le (-1)^{N-n-1}\left(\sigma_n(\boldsymbol{\alpha}, \boldsymbol{\beta})\varphi^2 + 2\sigma_{n+1}(\boldsymbol{\alpha}, \boldsymbol{\beta})\varphi + \sigma_{n+2}(\boldsymbol{\alpha}, \boldsymbol{\beta})\right) =: \xi_n(\varphi),$$
$$n = 0, \cdots, N-3. \tag{49}$$

The inequalities hold if and only if the discriminants of the quadratic functions $\xi_n(\varphi)$ are non-negative, $n = 0, \cdots, N-3$. The discriminant conditions are given by (16). □

Proof of Proposition 1. The theorem can be proven by applying Lemma

1 to every sub-interval $(a_t, b_t) = (z_t, z_{t+1})$, $t = 1, \cdots, T-1$.□

Proof of Proposition 2. Applying Lemma 2 to every sub-interval $(a_t, b_t) = (z_t, z_{t+1})$, $t = 1, \cdots, T-1$, we can establish the existence of a set of utility functions $f_t \in \mathcal{U}_N$ with $(\boldsymbol{\nabla f}_t(z_t), \boldsymbol{\nabla f}_t(z_{t+1})) = (\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_{t+1})$, $t = 1, \cdots, T-1$. Let

$$u\left(x\right) := \begin{cases} f_1\left(x\right) & x \in [z_1, z_2] \\ \vdots & \vdots \\ f_{T-1}(x) & x \in [z_{T-1}, z_T] \end{cases}. \tag{50}$$

The function is continuous at the sub-interval end points, or $\boldsymbol{\nabla f}_t(z_{t+1}) = \boldsymbol{\nabla f}_{t+1}(z_{t+1})$, $t = 1, \cdots, T-1$. Consequently, the local monotonicity properties of the local functions f_t, $t = 1, \cdots, T-1$, ensure global monotonicity and, therefore, $u \in \mathcal{U}_N$.

Furthermore, by construction,

$$\boldsymbol{\nabla u}_t(z_t) = \boldsymbol{\nabla f}_t(z_t) = \boldsymbol{\alpha}_t \tag{51}$$

for $t = 1, \cdots, T$.□

Appendix B: Optimization problems

In Section 7, we test TSD optimality of a given prospect. In this section, we present the CQP problem that we solved for that test. For the sake of generality, we specify the program for the general case of $M \geq 2$ and $1 \leq N \leq 4$. We combine Theorem 1 with Lemma 4 to obtain a finite system of convex quadratic constraints. To exclude zero solutions, we add the normalizing constraint $\sum_{s=1}^{S} p_{i,s}^{*} \alpha_{1,s} = 1$, following [PK13]. The objective is to minimize the largest feasible improvement in expected utility across all prospects, again following [PK13]. The resulting optimization problem follows:

$$
\begin{gathered}
\min \theta \\
\sum_{s=1}^{S} \alpha_{0,s}\left(p_{i,s}^{*}-p_{j,s}^{*}\right)+\theta \geq 0,\ j=1,\cdots,M; \\
(-1)^{n+1}\alpha_{n,s} \geq 0,\ n=0,\cdots,N-1; s=1,\cdots,S; \\
(-1)^{N-n-1}\sigma_{n}\left(\boldsymbol{\alpha}_{s},\boldsymbol{\alpha}_{s+1}\right) \geq 0,\ n=0,\cdots,N-1; s=1,\cdots,S-1; \\
\sigma_{n+1}(\boldsymbol{\alpha}_{s},\boldsymbol{\alpha}_{s+1})^{2}-\sigma_{n}(\boldsymbol{\alpha}_{s},\boldsymbol{\alpha}_{s+1})\sigma_{n+2}(\boldsymbol{\alpha}_{s},\boldsymbol{\alpha}_{s+1}) \leq 0,\ n=0,\cdots,N-3; \\
\sum_{s=1}^{S} p_{i,s}^{*}\alpha_{1,s}=1.
\end{gathered} \tag{52}
$$

Recall that $\sigma_n(\boldsymbol{\alpha}_s, \boldsymbol{\alpha}_{s+1})$ is a linear expression of $\boldsymbol{\alpha}_s$ and $\boldsymbol{\alpha}_{s+1}$; see (7) and (9). We end up with a CQP problem with model variables θ and $\alpha_{n,s}$, $n = 0, \cdots, N-1$; $s = 1, \cdots, S$. For $N = 1, 2$, the quadratic constraints vanish and we are left with an LP problem which is equivalent to the corresponding optimality problem in [PK13]. It follows from Theorem 1, that non-optimality occurs if and only if $\theta^* > 0$.

In Section 9, we test NSD efficiency of a market index ($2 \leq N \leq 5$). We now discuss the associated optimization problem. We combine Theorem 2 with Lemma 4 to obtain a finite system of convex quadratic constraints. To exclude zero solution, we add the normalizing constraint $\sum_{r=1}^{R} p_r \alpha_{1,r} = 1$. The objective is to minimize the largest error across all prospects. The resulting optimization problem follows:

$$
\begin{gathered}
\min \theta \\
\sum_{r=1}^{R} p_{r}\alpha_{1,r}\left(x_{r}^{*}-x_{j,r}\right)+\theta \geq 0,\ j=1,\cdots,M; \\
(-1)^{n+1}\alpha_{n,r} \geq 0,\ n=1,\cdots,N-1; r=1,\cdots,R; \\
(-1)^{N-n-1}\varsigma_{n}\left(\boldsymbol{\alpha}_{r},\boldsymbol{\alpha}_{r+1}\right) \geq 0,\ n=1,\cdots,N-1; r=1,\cdots,R-1; \\
\varsigma_{n+1}(\boldsymbol{\alpha}_{r},\boldsymbol{\alpha}_{r+1})^{2}-\varsigma_{n}(\boldsymbol{\alpha}_{r},\boldsymbol{\alpha}_{r+1})\varsigma_{n+2}(\boldsymbol{\alpha}_{r},\boldsymbol{\alpha}_{r+1}) \leq 0,\ n=1,\cdots,N-3; \\
\sum_{r=1}^{R} p_{r}\alpha_{1,r}=1.
\end{gathered} \tag{53}
$$

We arrive at a CQP problem with model variables θ and $\alpha_{n,r}$, $n = 1, \cdots, N-1$; $r = 1, \cdots, R$. For $N = 2, 3$, the quadratic constraints vanish and we find an LP problem that is equivalent to the corresponding efficiency problem in [PK13]. It follows from Theorem 2, that inefficiency occurs if and only if $\theta^* > 0$.

References

[AD15] Armbruster, B and E Delage, 2015, Decision Making under Uncertainty when Preference Information is Incomplete, *Management Science* 61, 111-128

[BP97] Atkinson, A, 1970, The measurement of inequality, Journal of Economic Theory 2, 244–263.

[Bawa75] Basso, A and P Pianca. 1997, Decreasing Absolute Risk Aversion And Option Pricing, *Management Science* 43, 206–216.

[BBRS85] Bawa, VS, 1975, Optimal Rules for Ordering Uncertain Prospects, *Journal of Financial Economics* 2, 95-121.

[BBRS85] Bawa, VS, JN Bodurtha Jr., MR Rao and HL Suri, 1985, On Determination of Stochastic Dominance Optimal Sets, *Journal of Finance* 40, 417-431.

[BLR79] Bawa, VS, EB Lindenberg and LC Rafsky, 1979, An Efficient Algorithm to Determine Stochastic Dominance Admissible Sets, *Management Science* 25, 609-622.

[CP96] Caballe, J and A Pomansky, 1996, Mixed risk aversion, *Journal of Economic Theory* 71, 485–513.

[DR03] Dentcheva, D and A Ruszczynski, 2003, Optimization with Stochastic Dominance Constraints, *SIAM Journal on Optimization* 14, 548-566.

[DK14] Dupacová J and M Kopa, 2014, Robustness of optimal portfolios under risk and stochastic dominance constraints, *European Journal of Operational Research* 234, 434-441.

[DI82] Dybvig, PH and JE Ingersoll, Jr., 1982, Mean-Variance Theory in Complete Markets, *The Journal of Business* 55, 233-251.

[EFR16] Eeckhoudt, L, AM Fiori, E Rosazza Gianin, 2016, Loss-averse preferences and portfolio choices: An extension, *European Journal of Operations Research* 249, 224-230.

[ES06] Eeckhoudt L and H Schlesinger, 2006, Putting risk in its proper place, *American Economic Review* 96, 280–289.

[Fang14] Fang, Y, 2014, Higher Order Stochastic Dominance and Aggregate Investor Preferences, SSRN working paper (August 18, 2014), http://ssrn.com/abstract=2708293.

[Fish74] Fishburn, PC, 1974, Convex stochastic dominance with continuous distribution functions, *Journal of Economic Theory* 7, 143-158.

[HS88] Hadar, J and TK Seo, 1988. Asset proportions in optimal portfolios, *The Review of Economic Studies* 55 (3), 459–468.

[HR69] Hadar, J and WR Russell, 1969, Rules for Ordering Uncertain Prospects, *American Economic Review* 59, 2–34.

[HH96] Hall, P and JL Horowitz, 1996, Bootstrap critical values for tests based on generalized-method-of-moments estimators, *Econometrica* 64, 891-916.

[HL69] Hanoch, G and H Levy, 1969, The Efficiency Analysis of Choices Involving Risk, *Review of Economic Studies* 36, 335–346.

[HHM14] Hu J, T Homem-de Mello and S Mehrotra, 2014, Stochastically weighted stochastic dominance concepts with an application in capital budgeting, *European Journal of Operational Research* 232, 572-583.

[KP09] Kopa, M and Th Post, 2009, A portfolio optimality test based on the first-order stochastic dominance criterion, *Journal of Financial and Quantitative Analysis* 44, 1103–1124.

[KP15] Kopa, M and Th Post, 2015, A General Test for SSD Portfolio Efficiency, *OR Spectrum* 37, 703-734.

[Kuos04] Kuosmanen, T, 2004, Efficient diversification according to stochastic dominance criteria, *Management Science* 50, 1390-1406.

[LL02] Leshno, M, Levy, H, 2002, Preferred by "all" and preferred by "most" decision makers: Almost stochastic dominance, *Management Science* 48, 1074-1085.

[Liz12a] Lizyayev A, 2012, Stochastic dominance: convexity and some efficiency tests, *Int. J. Theor. Appl. Financ.* 15, 1–19.

[Liz12b] Lizyayev A, 2012, Stochastic Dominance Efficiency Analysis of Diversified Portfolios: Classification, Comparison and Refinements, *Annals of Operational Research* 196, 391-410.

[LR12] Lizyayev A, and A Ruszczynski, 2012, Tractable almost stochastic dominance, *European Journal of Operations Research* 218, 448–455.

[Long16] Longarela, IR, 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.

[MXF12] Meskarian R, H Xu, J Fliege, 2012, Numerical methods for stochastic programs with second order dominance constraints with applications to portfolio optimization, *European Journal of Operational Research* 216, 376-385.

[MSTW16] Meyer, J, 1977, Second degree stochastic dominance with respect to a function, *International Economic Review* 18, 477-487.

[MSTW16] Müller, A, M Scarsini, I Tsetlin and RL Winkler, 2016, Between First and Second-Order Stochastic Dominance, forthcoming in *Management Science*.

[Pod14] Podinovski VV, 2014, Decision making under uncertainty with unknown utility function and rank-ordered probabilities, *European Journal of Operational Research* 239, 537-541.

[Post03] Post, Th, 2003, Empirical Tests for Stochastic Dominance Efficiency, *Journal of Finance* 58, 1905-1932.

[Post16] Post, Th, 2016, Standard Stochastic Dominance, *European Journal of Operational Research* 248, 1009-1020.

[PFK15] Post, Th, Y Fang, and M Kopa, 2015, Linear Tests for DARA Stochastic Dominance, *Management Science* 61, 1615-1629.

[PK13] Post, Th and M Kopa, 2013, General Linear Formulations of Stochastic Dominance Criteria, *European Journal of Operational Research* 230, 321-332.

[PP16] Post, Th and V Poti, 2016, Portfolio Analysis using Stochastic Dominance, Relative Entropy and Empirical Likelihood, forthcoming in *Management Science*.

[PK16] Post, Th and M Kopa, 2016, Portfolio Choice based on Third-degree Stochastic Dominance, forthcoming in *Management Science*.

[PV07] 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.

[PWF73] Porter, RB, JR Wart and DL Ferguson, 1973, Efficient Algorithms for Conducting Stochastic Dominance Tests on Large Numbers of Portfolios, The Journal of Financial and Quantitative Analysis 8, 71-81.

[QS62] Quirk, JP and R Saposnik, 1962, Admissibility and Measurable Utility Functions, *Review of Economics Studies* 29, 140-146.

[RMZ13] Roman D, G Mitra, V Zverovich, 2013, Enhanced indexation based on second order stochastic dominance, *European Journal of Operational Research* 228, 273-281.

[RS70] Rothschild, M and JE Stiglitz, 1970, Increasing Risk: I. A Definition, *Journal of Economic Theory*, 225-243.

[RS89] Russell, WR and TK Seo, 1989, Representative sets for stochastic dominance rules, In: Fomby, TB, Seo, TK (Eds.), *Studies in the Economics of Uncertainty: In Honor of Josef Hadar*. Springer Verlag, New York, 59–76.

[RDM06] Roman, D, K Darby-Dowman and G Mitra, 2006, Portfolio construction based on stochastic dominance and target return distributions, *Mathematical Programming* 108, 541-569.

[SY94] Scaillet, O and N Topaloglou, 2010, Testing for stochastic dominance efficiency, *Journal of Business and Economic Statistics* 28, 169–180.

[ST94] Shalit, H and S Yitzhaki, 1994, Marginal Conditional Stochastic Dominance, *Management Science* 40, 670–684.

[THS13] Tzeng, LY, RJ Huang and P-T Shih, 2013, Revisiting Almost Second-Degree Stochastic Dominance, *Management Science* 59, 1250-1254.

[Vick75] Vickson, RG, 1975, Stochastic Dominance Tests for Decreasing Absolute Risk Aversion. I. Discrete Random Variables, *Management Science* 21, 1438–1446.

[Whit70] Whitmore, GA, 1970, Third-degree Stochastic Dominance, *American Economic Review* 60, 457–59.