Fractional Stochastic Dominance in Rank-dependent Utility and Cumulative Prospect Theory*

Tiantian Mao† Ruodu Wang‡

August 29, 2022

Abstract

Two notions of fractional stochastic dominance (SD) were recently proposed by Müller et al. (2017) and Huang et al. (2020) based on mean-reducing spreads and the coefficient of absolute risk aversion, respectively. We formulate a general class of fractional SD generated by a convex transform, which includes those built from absolute or relative risk aversion as special cases, and this serves as a convenient technical tool for construction of new notions of fractional SD. We obtain equivalent conditions for a preference modelled by rank-dependent utility or cumulative prospect theory to be consistent with each notion of fractional SD. Furthermore, we provide an empirical estimator for the parameters in fractional SD relationships, and we illustrate this with a financial data analysis.

Keywords: stochastic dominance, risk aversion, risk measures, rank-dependent utility, cumulative prospect theory

Declarations of interest: none.

1 Introduction

Two notions of fractional stochastic dominance (SD) were recently proposed by Müller et al. (2017) and Huang et al. (2020), respectively. Fractional SD was introduced because the classic

*We thank the Editor, three anonymous referees, and Lin Zhao for many helpful comments. An extended version of this paper previously circulated under the title "A critical comparison of three notions of fractional stochastic dominance". T. Mao acknowledges financial support from National Natural Science Foundation of China (Nos. 71871208, 71921001, 71671176). R. Wang acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC RGPIN-2018-03823, RGPAS-2018-522590).

†Department of Statistics and Finance, School of Management, University of Science and Technology of China, China. Email: tmao@ustc.edu.cn

‡Department of Statistics and Actuarial Science, University of Waterloo, Canada. Email: wang@uwaterloo.ca.

notions of first-order SD (FSD) and second-order SD (SSD) are too often coarse and they could not capture, e.g., local convexity of a utility function in the expected utility (EU) model. Studies of stochastic dominance help to analyze decisions for a class of heterogeneous decision makers sharing some similarity in their risk attitude, without specifying the preference of a particular decision maker. The first notion of Müller et al. (2017) is based on γ-spread for $\gamma \in [0,1]$,[1] and will henceforth be referred to as $(1+\gamma)_S$-SD; see also Müller et al. (2021). The second notion of Huang et al. (2020) is based on the Arrow-Pratt coefficient of absolute risk aversion, and will be referred to as $(1+c)_A$-SD for $c \in [0,1]$. Precise definitions are put in Section 2. For comprehensive discussions on the relevance of these notions, we refer to Müller et al. (2017) and Huang et al. (2020).

Risk aversion has been a critical concept in decision making since Pratt (1964) and Arrow (1974). Various notions of risk aversion were proposed, observed and tested from empirical studies by Harrison (1986), Tversky and Kahneman (1992), Kimball (1993), Rabin (2000), Rabin and Thaler (2001) and Schmidt and Traub (2002), amongst others. As the most classic notion, a preference relation is strongly risk averse (Hadar and Russell (1969), Hanoch and Levy (1969) and Rothschild and Stiglitz (1970)) if it is monotone in SSD. Since fractional SD bridges SSD and FSD, monotonicity in fractional SD can be seen as a property of fractional risk aversion.

Our main contribution is a characterization of monotonicity in fractional SD for the behavioral decision models of rank-dependent utility (RDU) of Quiggin (1982)[2] and cumulative prospect theory (CPT) of Tversky and Kahneman (1992). The considered notions of fractional SD include $(1+\gamma)_S$-SD, $(1+c)_A$-SD and the latter's analogue based on relative risk aversion. Both RDU and CPT are generalizations of the EU model and the dual utility theory (DT) of Yaari (1987). Although RDU can be seen as a special case of CPT, conditions for fractional SD in RDU are more mathematically concise and economically interpretable, and hence we will present RDU and CPT results separately in Section 3.

To explain our motivation for studying fractional SD in behavioral models, we look again at FSD and SSD, which are limits of fractional SD. FSD and SSD were traditionally formulated based on EU, although these properties are model free. For instance, SSD can be equivalently formulated via mean-preserving spread (Rothschild and Stiglitz (1970)), conditional expectations (Strassen (1965)), dual utility (Yaari (1987)), and aversion to positive dependence (Wang and Wu (2020)); the case of FSD is similar. Likewise, the notions of fractional SD are suitable for study beyond EU, and in particular, we are interested in their implication for the popular descriptive decision models

[1] Müller et al. (2017) used the term "γ-transfer". We use "γ-spread" (i.e., the inverse of a γ-transfer) as this notion closely resembles the mean-preserving spreads of Rothschild and Stiglitz (1970). Indeed, γ-spreads are mean-reducing spreads.

[2] As shown by Chew et al. (1987), a necessary condition for monotonicity in SSD for RDU is a concave probability perception function. The functons originally used by Quiggin (1982) is not concave.

of RDU and CPT.

Equivalent characterization of strong risk aversion in different decision models has been studied by Pratt (1964) and Arrow (1974) for EU, Yaari (1987) for DT, Chew et al. (1987) for RDU, and Schmidt and Zank (2008) for CPT. Our results generalize the above results to several formulations of fractional SD, including $(1+\gamma)_{\mathrm{S}}$-SD and $(1+c)_{\mathrm{A}}$-SD. Moreover, characterization results are obtained for a class of fractional SD connected to SSD via a transform v, called v-SD, which includes $(1+c)_{\mathrm{A}}$-SD as special cases.

As we will see in Section 2, $(1+c)_{\mathrm{A}}$-SD is closely related to an exponential transformation. More precisely, X is dominated by Y in $(1+c)_{\mathrm{A}}$-SD if and only if $e^{\lambda X}$ is dominated by $e^{\lambda Y}$ in SSD, where $\lambda = 1/c - 1$. Therefore, many results and convenient properties of SSD can be translated to those of $(1+c)_{\mathrm{A}}$-SD. A negative result (Proposition 2) implies that there does not exist a risk transform or probability distortion such that $(1+\gamma)_{\mathrm{S}}$-SD can be associated with SSD. To be more precise, $(1+\gamma)_{\mathrm{S}}$-SD between X and Y cannot be equivalently described by SSD between $v(X)$ and $v(Y)$ for any transform v, and this remains so if we further allows for probability distortions as in RDU. This nonexistence illustrates that the mathematical basis of $(1+\gamma)_{\mathrm{S}}$-SD is fundamentally different from SSD and $(1+c)_{\mathrm{A}}$-SD, which also explains why technical results such as a characterization in RDU and CPT are much more complicated for $(1+\gamma)_{\mathrm{S}}$-SD than that for $(1+c)_{\mathrm{A}}$-SD.

To better understand applications of fractional SD in behavioral decision models outside EU, we proceed to study further technical properties of fractional SD in Section 4, empirical estimators of parameters of fractional SD between two distributions in Section 5, and a real-data analysis in Section 6. These additional results illustrate what we can analyze when fractional SD is brought outside EU.

All proofs are relegated to Appendix A. Some simulation results for estimating the parameters γ and c, complementing Section 6, are put in Appendix B.

2 Notions of fractional stochastic dominance

Similarly to Müller et al. (2017) and Huang et al. (2020), we use a unified notation $[a, b]$ where $a < b$ to denote an interval containing the support of the random variables, which encompasses $[a, \infty)$ if $b = \infty$, $(-\infty, b]$ if $a = -\infty$, and $(-\infty, \infty)$ if both. All utility functions are elements of $\mathscr{U} = \{u : [a, b] \to \mathbb{R} \mid u \text{ is increasing and twice differentiable}\}$.[3] In all the statements below, X and Y are arbitrary random variables taking values in $[a, b]$, and inequalities on expectations are meant to hold when both sides are well defined.

[3] In this paper, the term "increasing" is in the non-strict sense. The differentiability condition can be safely relaxed without changing the corresponding fractional SD defined in this paper; see e.g., Müller et al. (2017) and Huang et al. (2020).

The first notion of fractional SD is $(1+\gamma)_{\mathrm{S}}$-SD of Müller et al. (2017). Throughout, random variables X and Y represent random payoffs (prospects) to be compared via notions of SD. For $\gamma \in [0,1]$, define the set $\mathscr{U}_\gamma^{\mathrm{S}}$ of utilities by

$$\mathscr{U}_\gamma^{\mathrm{S}} = \{u \in \mathscr{U} \mid \gamma u'(y) \leqslant u'(x) \text{ for all } a \leqslant x \leqslant y \leqslant b\}.$$

Definition 1 (Müller et al. (2017))**.** For fixed $\gamma \in [0,1]$, Y dominates X by $(1+\gamma)_{\mathrm{S}}$-SD, denoted by $X \leqslant_\gamma^{\mathrm{S}} Y$, if $\mathbb{E}[u(X)] \leqslant \mathbb{E}[u(Y)]$ for all $u \in \mathscr{U}_\gamma^{\mathrm{S}}$. The subscript/superscript "S" here represents that the notion is defined via γ-spread.

Huang et al. (2020) proposed $(1+c)_{\mathrm{A}}$-SD based on utility functions with a lower bound on the Arrow-Pratt coefficient of absolute risk aversion. For a twice continuously differentiable utility function u, the coefficient of absolute risk aversion is defined as

$$\rho_u^{\mathrm{A}}(x) = -\frac{u''(x)}{u'(x)}.$$

For $c \in [0,1]$, define the set $\mathscr{U}_c^{\mathrm{A}}$ of utilities by

$$\mathscr{U}_c^{\mathrm{A}} = \left\{u \in \mathscr{U} \mid u'(x) > 0 \text{ and } \rho_u^{\mathrm{A}}(x) \geqslant -\frac{1}{c} + 1 \text{ for all } x \in [a,b]\right\}.$$

Definition 2 (Huang et al. (2020))**.** For fixed $c \in [0,1]$, Y dominates X by $(1+c)_{\mathrm{A}}$-SD, denoted by $X \leqslant_c^{\mathrm{A}} Y$, if $\mathbb{E}[u(X)] \leqslant \mathbb{E}[u(Y)]$ for all $u \in \mathscr{U}_c^{\mathrm{A}}$. The subscript/superscript "A" here represents that the notion is defined via the coefficient of absolute risk aversion.

Replacing the coefficient of absolute risk aversion ρ_u^{A} by the coefficient of relative risk aversion

$$\rho_u^{\mathrm{R}}(x) = -\frac{x\, u''(x)}{u'(x)},$$

we arrive at another notion of fractional SD, which is briefly discussed in the concluding remarks of Huang et al. (2020). Because of its connection to relative risk aversion, this notion is only defined for nonnegative random variables. For $r \in [0,1]$, define the set $\mathscr{U}_r^{\mathrm{R}}$ of utilities by

$$\mathscr{U}_r^{\mathrm{R}} = \left\{u \in \mathscr{U} \mid u'(x) > 0 \text{ and } \rho_u^{\mathrm{R}}(x) \geqslant -\frac{1}{r} + 1 \text{ for all } x \in [a,b]\right\}.$$

Definition 3. For fixed $r \in [0,1]$, $0 \leqslant a < b \leqslant \infty$, Y dominates X by $(1+r)_{\mathrm{R}}$-SD, denoted by $X \leqslant_r^{\mathrm{R}} Y$, if $\mathbb{E}[u(X)] \leqslant \mathbb{E}[u(Y)]$ for all $u \in \mathscr{U}_r^{\mathrm{R}}$. The subscript/superscript "R" here represents that the notion is defined via the coefficient of relative risk aversion.

By definition, 1_{S}-SD, 1_{A}-SD and 1_{R}-SD are all equivalent to the classic FSD (denoted by $\leqslant_{\mathrm{FSD}}$)

and 2_S-SD, 2_A-SD and 2_R-SD are all equivalent to the classic SSD (denoted by $\leqslant_{SSD}$).

The notions of $(1+c)_A$-SD and $(1+r)_R$-SD can be unified under the umbrella of v-SD, which was first studied by Meyer (1977). Let $v \in \mathscr{U}$ be a convex function with $v' > 0$. Define the set $\mathscr{U}_v$ of utility functions by

$$\mathscr{U}_v = \left\{u \in \mathscr{U} \mid u'(x) > 0,\ \rho_u^A(x) \geqslant \rho_v^A(x),\ x \in [a,b]\right\}. \tag{1}$$

Taking $v(x) = e^{(1/c-1)x}$ leads to $\rho_v^A(x) = -(1/c-1)$ and $\mathscr{U}_v = \mathscr{U}_c^A$. Similarly, taking $v(x) = x^{1/r}$ gives $\rho_v^A(x) = -(1/r-1)/x$ such that $\mathscr{U}_v = \mathscr{U}_r^R$. Therefore, $(1+c)_A$-SD and $(1+r)_R$-SD are both special cases of the fractional SD generated by $\mathscr{U}_v$ for a specific function v.

Definition 4. For a convex function $v \in \mathscr{U}$ with $v' > 0$, Y dominates X by v-SD, denoted by $X \leqslant_v Y$, if $\mathbb{E}[u(X)] \leqslant \mathbb{E}[u(Y)]$ for all $u \in \mathscr{U}_v$.

Although v-SD is defined with EU conditions, it can be used together with decision models other than EU, just like any partial order between random variables. For instance, SSD and FSD are often defined with EU conditions, and they are applied to a wide range of decision models. For alternative formulations of $(1+\gamma)_S$-SD and $(1+c)_A$-SD without using EU conditions, see Müller et al. (2017) and Huang et al. (2020).

As shown by Pratt (1964, Theorem 1), $\rho_u^A \geqslant \rho_v^A$ is equivalent (up to differentiability) to $u = w \circ v$ for some increasing concave function w (here, $\circ$ is the composition of two functions), and hence $\mathscr{U}_v$ can be safely replaced by

$$\mathscr{U}_v^* = \{u \in \mathscr{U} \mid u(x) = w(v(x)),\ x \in [a,b] \text{ for some increasing concave function } w\}. \tag{2}$$

This reformulation immediately allows us to translate between v-SD and SSD by noting that $\mathbb{E}[w(v(X))] \leqslant \mathbb{E}[w(v(Y))]$ for all strictly increasing concave w is equivalent to $v(X) \leqslant_{SSD} v(Y)$.

Proposition 1 (Meyer (1977))**.** *Take any convex function $v \in \mathscr{U}$ with $v' > 0$. For all X and Y, $X \leqslant_v Y$ if and only if $v(X) \leqslant_{SSD} v(Y)$.*

Proposition 1 leads to the following SSD-based formulation of $(1+c)_A$-SD and $(1+r)_R$-SD.

Corollary 1. *For any X, Y and $c, r \in (0,1)$,*

$$X \leqslant_c^A Y \iff e^{(1/c-1)X} \leqslant_{SSD} e^{(1/c-1)Y} \quad \text{and} \quad X \leqslant_r^R Y \iff X^{1/r} \leqslant_{SSD} Y^{1/r}.$$

Since $(1+c)_A$-SD and $(1+r)_R$-SD, as well as v-SD in general, are both connected to SSD via a transformation, an immediate question that emerges concerns the existence of a similar translation between $(1+\gamma)_{\mathrm{S}}$-SD and SSD. The answer is negative even if we allow for both shape transforms and probability distortions (i.e., $F \mapsto h \circ F \circ v$ for some h and v), which are two fundamental distributional transforms characterized by Liu et al. (2021). This suggests some fundamental difference between $(1+\gamma)_{\mathrm{S}}$-SD and v-SD. A precise statement of this negative result is given below.

Proposition 2. *For any $\gamma < 1$, there do not exist $v : \mathbb{R} \to \mathbb{R}$ and $h : [0,1] \to [0,1]$ such that for all X, Y taking values on $[a,b]$,*

$$X \leqslant_{\gamma}^{\mathrm{S}} Y \iff (v(X))_h \leqslant_{\mathrm{SSD}} (v(Y))_h. \tag{3}$$

where Z_h represents a random variable having the distribution function $h \circ H$ and H is the distribution of Z.

For each given function $v > 0$, we can build a continuum of fractional SD via (v^t)-SD indexed by $t \in (t_0, \infty)$, that is, to consider the set of utility functions (where v^t means v raised to the power of t)

$$\mathscr{U}_{v,t} = \left\{ u \in \mathscr{U} \mid u(x) = w\left(v^t(x)\right),\ x \in [a,b],\ w \text{ is an increasing concave function} \right\},$$

and t_0 is such that v^t is convex for $t > t_0$. To obtain the class of $(1+c)_{\mathrm{A}}$-SD, we choose $t_0 = 0$ and $v(x) = e^x$. To obtain the class of $(1+r)_{\mathrm{R}}$-SD, we choose $t_0 = 1$ and $v(x) = x$. The property of (v^t)-SD in this continuum becomes stronger as t increases (Remark 1 below), and one can employ a decreasing transform from (t_0, ∞) to $(0,1)$ if it is desirable to index the class by a parameter in $(0,1)$. Because of the economic relevance of $(1+c)_{\mathrm{A}}$-SD and $(1+r)_{\mathrm{R}}$-SD, we will be primarily interested in these two special cases of v-SD, although all our results in Section 3 are presented for general v-SD.

Remark 1. Monotonicity of v-SD with respect to the coefficient of absolute risk aversion of v is straightforward. That is, if $\rho_{v_1}^{\mathrm{A}} \leqslant \rho_{v_2}^{\mathrm{A}}$, then $\geqslant_{v_1}$ is stronger than $\geqslant_{v_2}$. In particular, $\geqslant_{v^r}$ is stronger than $\geqslant_{v^s}$ for $t \geqslant s$.

One may naturally wonder whether the class of v^t-SD recovers FSD and SSD as its limiting cases in the sense that $\rho_{v^t}^{\mathrm{A}}(x) \to -\infty$ as $t \to \infty$, and $\rho_{v^t}^{\mathrm{A}}(x) \to 0$ as $t \downarrow t_0$ for all $x \in [a,b]$, respectively. The next proposition shows that this is true only in the two cases of $(1+c)_{\mathrm{A}}$-SD and $(1+r)_{\mathrm{R}}$-SD.

Proposition 3. *For a positive and twice differentiable function v with $v' > 0$ and $t_0 \geqslant 0$, the family of (v^t)-SD always recovers FSD as $t \to \infty$. The family of (v^t)-SD covers SSD as $t \downarrow t_0$ if and only if (v^t)-SD is one of $(1+c)_{\mathrm{A}}$-SD or $(1+r)_{\mathrm{R}}$-SD for some $c, r \in [0,1]$.*

As a consequence of Proposition 3, the function v which guarantees that the family of (v^t)-SD recovers FSD and SSD is a linear transform of a power or exponential function.

3 Fractional stochastic dominance in RDU and CPT

Below we present our results on fractional SD in the popular behavioral decision models of RDU and CPT. Since CPT can be seen as a generalization of RDU, we only need to prove the characterization results for CPT; nevertheless, since the formulas and conditions in RDU are more concise and easier to interpret, we first present the RDU results, followed by the CPT results.

In the RDU model, each decision maker is characterized by a utility function $u : \mathbb{R} \to \mathbb{R}$ and a probability perception function $h : [0, 1] \to [0, 1]$ with $h(0) = 0$, $h(1) = 1$. Both u and h are assumed to be increasing and continuous. For given u and h, the rank-dependent utility of a prospect X is defined by

$$V_{u,h}(X) = \int_{\mathbb{R}} u(x)\,\mathrm{d}h(F(x)), \tag{4}$$

where F is the distribution function of X. An RDU(u, h) preference $\precsim$ is given by

$$X \precsim Y \iff V_{u,h}(X) \leqslant V_{u,h}(Y). \tag{5}$$

A preference $\precsim$ (a total preorder on random variables supported in $[a, b]$) is *monotone* in the partial order $\leqslant^{\mathrm{S}}_{\gamma}$ (or $\leqslant_v$) if $X \leqslant^{\mathrm{S}}_{\gamma} Y$ (or $X \leqslant_v$ Y) implies $X \precsim Y$. Strong risk aversion corresponds to monotonicity in SSD. As shown by Chew et al. (1987), an RDU(u, h) preference is strongly risk averse if and only if both u and h are concave. This statement is generalized to $(1 + \gamma)_{\mathrm{S}}$-SD and v-SD in the following theorem.

Theorem 1. *For a convex function $v \in \mathscr{U}$ with $v' > 0$ and $\gamma \in (0, 1)$, the following statements hold for any utility u function and probability perception function h.*

(i) An RDU(u, h) preference is monotone in $\leqslant_v$ if and only if $u(v^{-1}(y))$ is concave in $y \in \mathbb{R}$ and h is concave. In particular,

(a) for $c \in (0, 1)$, an RDU(u, h) preference is monotone in $\leqslant^{\mathrm{A}}_c$ if and only if $u(\frac{c}{1-c}\log y)$ is concave in $y \in \mathbb{R}_+$ and h is concave;

(b) for $r \in (0, 1)$, an RDU(u, h) preference is monotone in $\leqslant^{\mathrm{R}}_r$ if and only if $u(y^r)$ is concave in $y \in \mathbb{R}_+$ and h is concave.

(ii) An RDU(u,h) preference is monotone in $\leqslant_\gamma^{\mathrm{S}}$ if and only if $P_h \geqslant \gamma G_u^{[a,b]}$, where

$$P_h = \inf_{0\leqslant p_1<p_2\leqslant p_3<p_4\leqslant 1} \frac{h(p_2)-h(p_1)}{p_2-p_1} \Big/ \frac{h(p_4)-h(p_3)}{p_4-p_3}, \tag{6}$$

and

$$G_u^{[a,b]} = \sup_{a\leqslant x_1<x_2\leqslant x_3<x_4\leqslant b} \frac{u(x_4)-u(x_3)}{x_4-x_3} \Big/ \frac{u(x_2)-u(x_1)}{x_2-x_1}. \tag{7}$$

Taking $v(x)=x$ in (i) or $\gamma=1$ in (ii) yields the condition for strong risk aversion in RDU of Chew et al. (1987) that h and u are both concave. Note that $G_u^{[a,b]} \geqslant 1 \geqslant P_h$, $P_h=1$ if and only if h is concave, and $G_u^{[a,b]}=1$ if and only if u is concave on $[a,b]$.

The quantities $G_u^{[a,b]}$ and P_h are called the index of greediness and the index of pessimism, respectively, by Chateauneuf et al. (2005). Therefore, a simple interpretation of Theorem 1 (ii) is that monotonicity in $(1+\gamma)_S$-SD in RDU means a balance between greediness and pessimism. On the other hand, Theorem 1 (i) says that for $(1+c)_{\mathrm{A}}$-SD and $(1+r)_{\mathrm{R}}$-SD, the requirement on the probability perception function h is the same as SSD, and the utility function u needs to be "not too convex"; if u is twice differentiable, then this means it needs to be in $\mathscr{U}_c^{\mathrm{A}}$ or $\mathscr{U}_r^{\mathrm{R}}$.

If the utility function u is the identity, the RDU model reduces to Yaari's DT, denoted by DT(h).

(i) A DT(h) preference is monotone in v-SD if and only if h is concave. Hence, monotonicity in v-SD for any choice of v is equivalent to strong risk aversion in DT. This shows that DT is not able to distinguish, for instance, different values of the parameter c in $(1+c)_{\mathrm{A}}$-SD.

(ii) A DT(h) preference is monotone in $\leqslant_\gamma^{\mathrm{S}}$ if and only if $P_h \geqslant \gamma$, where P_h is defined by (6). This condition allows for some h that is not concave, and thus monotonicity in $\leqslant_\gamma^{\mathrm{S}}$ is genuinely different from strong risk aversion. Indeed, $\leqslant_\gamma^{\mathrm{S}}$ may be alternatively defined using DT instead of EU, and this is not the case for $(1+c)_{\mathrm{A}}$-SD.

As we see in Theorem 1, in order to determine whether an RDU preference is monotone in $\leqslant_v$, it suffices to check whether h is concave and $u\circ v^{-1}$ is concave. In case of $\leqslant_c^{\mathrm{A}}$, this corresponds to RDU with the utility functions in Huang et al. (2020) plus a concave h.

To determine whether an RDU preference is in $\leqslant_\gamma^{\mathrm{S}}$, we need to compute P_h and $G_u^{[a,b]}$. We present P_h and $G_u^{[a,b]}$ for some examples u and h.

Example 1 (Utility functions). (i) Consider the piece-wise linear utility function $u(x)=\alpha x_+ - \beta x_-$, $x\in\mathbb{R}$, where $\alpha,\beta>0$, and $a<0<b$. If $\alpha\leqslant\beta$ (risk-neutral or risk-averse), then $G_u^{[a,b]}=1$. If $\alpha>\beta$ (risk-seeking), then $G_u^{[a,b]}=\alpha/\beta$.

(ii) Consider the exponential utility function $u(x) = \frac{1}{\beta}(1 - e^{-\beta x})$, $x \in \mathbb{R}$, where $\beta \neq 0$. We have

$$G_u^{[a,b]} = \max\left\{1, e^{\beta(b-a)}\right\}.$$

Note that if $\beta > 0$ then u is concave, and if $\beta < 0$ then u is convex.

(iii) Consider the power utility function $u(x) = (w_0 + x)^\beta$, $x > -w_0$, where $\beta > 0$ and $w_0 \in \mathbb{R}$ represents the initial wealth of this decision-maker. Then for $a > -w_0$,

$$G_u^{[a,b]} = \max\left\{1, \left(\frac{w_0 + b}{w_0 + a}\right)^{\beta - 1}\right\}.$$

Note that if $\beta \leqslant 1$ then u is concave, and if $\beta \geqslant 1$ then u is convex.

(iv) Consider the utility function of Tversky and Kahneman (1992),

$$u(x) = x^\alpha \mathbb{1}_{\{x \geqslant 0\}} - \nu(-x)^\beta \mathbb{1}_{\{x<0\}}, \quad x \in \mathbb{R},$$

where $\alpha, \beta, \nu > 0$. The function u is an inverse-S-shaped utility function and $u'_+(0) = \infty$ if $0 < \beta < 1 < \alpha$. We exclude the case that $\alpha = \beta = 1$, which was discussed in (i). We have

$$G_u^{[a,b]} = \begin{cases} \max\{(b/a)^{\alpha-1}, 1\}, & 0 \leqslant a < b, \\ \max\{(b/a)^{\beta-1}, 1\}, & a < b < 0, \\ \infty, & \text{otherwise.} \end{cases}$$

Example 2 (Probability perception functions). (i) Consider the hyperbolic perception function of Chateauneuf et al. (2005), defined by $h(p) = p/(p + \nu(1-p))$, $p \in [0,1]$, for some $\nu > 1$. The function h is convex, and we can calculate $P_h = h'(0)/h'(1) = 1/\nu^2$.

(ii) Consider the power perception function $h(p) = p^\alpha$, $p \in [0,1]$, where $\alpha > 0$. We can calculate $P^{[a,b]} = \mathbb{1}_{\{\alpha \in (0,1]\}}$.

(iii) Consider the inverse-S-shaped probability perception function introduced by Tversky and Kahneman (1992),

$$h(p) = \frac{p^\delta}{(p^\delta + (1-p)^\delta)^{1/\delta}}, \quad p \in [0,1],$$

where $\delta \in (0,1)$. One can calculate $P_h = 0$.

Combining some u and h in Examples 1 and 2 we obtain some RDU preferences monotone in $\leqslant_\gamma^{\mathrm{S}}$ but not in SSD. An example is provided below.

Example 3. Fix $b > a > 0$. Consider the power utility function $u(x) = x^\beta$ and the hyperbolic perception function $h(p) = p/(p + \nu(1-p))$ with $\beta, \nu > 1$. Examples 1 and 2 yield that $P_h = 1/\nu^2$ and $G_u^{[a,b]} = (b/a)^{\beta-1}$, and we let $\gamma_0 := P_h/G_u^{[a,b]} = \nu^{-2}(b/a)^{1-\beta}$. Then, by Theorem 1, the RDU(u, h) preference is monotone in $\leqslant_\gamma^{\mathrm{S}}$ with $\gamma \leqslant \gamma_0$ but monotonicity fails to hold for $\gamma > \gamma_0$. For instance, if $\nu = \beta = 3/2$ and $b/a = 4$, then $\gamma_0 = 2/9$. Also note that $\gamma_0 \uparrow 1$ as $\nu, \beta \downarrow 1$. On the other than, this RDU(u, h) preference cannot be monotone in $\leqslant_c^{\mathrm{A}}$ or $\leqslant_r^{\mathrm{R}}$ for any $c, r \in (0, 1]$, since h is not concave.

Next, we turn to the cumulative prospect theory (CPT) of Tversky and Kahneman (1992). In the CPT model, for an increasing continuous function $u : \mathbb{R} \to \mathbb{R}$, two distortion functions h_1, h_2, and a risk X, the expected loss/utility based on CPT is defined as

$$V_{u,h_1,h_2}(X) = \int_{-\infty}^{0} u(x)\,\mathrm{d}h_1(F(x)) + \int_0^\infty u(x)\,\mathrm{d}h_2(F(x)), \tag{8}$$

where F is the distribution function of X. A CPT(u, h_1, h_2) preference $\precsim$ is given by

$$X \precsim Y \iff V_{u,h_1,h_2}(X) \leqslant V_{u,h_1,h_2}(Y). \tag{9}$$

If $h_1 = h_2 = h$, then CPT(u, h_1, h_2) is an RDU preference. Moreover, we will assume $a < 0 < b$; that is, the reference point 0 of CPT is contained in $[a, b]$; otherwise (8) again reduces to (4).

Theorem 2. *(i) A CPT(u, h_1, h_2) preference is monotone in $\leqslant_v$ if and only if $u(v^{-1}(y))$ is both concave in $y \in [0, v(0)]$ and $y \in [v(0), \infty)$, h_1, h_2 are concave, and they satisfy*

$$\frac{u'_+(0)}{u'_-(0)} \leqslant \inf_{p\in(0,1)} \frac{(h_1)'_-(p)}{(h_2)'_+(p)}. \tag{10}$$

(ii) A CPT(u, h_1, h_2) preference is monotone in $\leqslant_\gamma^{\mathrm{S}}$ if and only if $\gamma G_u^{[a,0]} \leqslant P_{h_1}$, $\gamma G_u^{[0,b]} \leqslant P_{h_2}$ and $\gamma G_u^{*[a,b]} \leqslant P_{h_1,h_2}$, where P_h and $G_u^{[a,b]}$ are defined by (6) and (7), respectively, and

$$P_{h_1,h_2} = \inf_{0\leqslant p_1<p_2\leqslant p_3<p_4\leqslant 1} \frac{h_1(p_2) - h_1(p_1)}{p_2 - p_1} \Big/ \frac{h_2(p_4) - h_2(p_3)}{p_4 - p_3}, \tag{11}$$

$$G_u^{*[a,b]} = \sup_{a\leqslant x_1<x_2\leqslant 0\leqslant x_3<x_4\leqslant b} \frac{u(x_4) - u(x_3)}{x_4 - x_3} \Big/ \frac{u(x_2) - u(x_1)}{x_2 - x_1}. \tag{12}$$

Theorem 2 generalizes the characterization result of risk aversion in CPT by Schmidt and Zank (2008). An empirically observed feature in CPT is loss aversion, which means $u'(-x) \geqslant u'(x)$ for all $x > 0$ where the derivatives exist; see e.g., Baucells and Heukamp (2006). Note that the limit of $u'(x)/u'(-x)$ as $x \downarrow 0$, if it exists, yields the left-hand side of (10). Similarly to the interpretation of Theorem 1, monotonicity in $(1+\gamma)_S$-SD in CPT means a subtle balance between greediness and pessimism for both the positive part $[0,b]$ and the negative part $[a,0]$, and monotonicity in v-SD implies that u is "not too convex" and h_1 and h_2 are concave.

4 Operational properties and applicability for specific models

We proceed to compare the three notions of fractional SD in terms of operational properties. These properties further illustrate the differences between these notions of fractional SD in behavioral models beyond EU.

We first state three invariance properties, where (i) and (ii) are shown by Müller et al. (2017) and Huang et al. (2020). For any prospects X and Y and $\gamma, c, r \in [0,1]$,

(i) $X \leqslant_\gamma^{\mathrm{S}} Y$ if and only if $\alpha X+\beta \leqslant_\gamma^{\mathrm{S}} \alpha Y+\beta$ for all $\alpha \geqslant 0$ and $\beta \in \mathbb{R}$;

(ii) $X \leqslant_c^{\mathrm{A}} Y$ if and only if $\alpha X+\beta \leqslant_{\alpha c/(1-c+\alpha c)}^{\mathrm{A}} \alpha Y+\beta$ for all $\alpha \geqslant 0$ and $\beta \in \mathbb{R}$.

(iii) $X \leqslant_r^{\mathrm{R}} Y$ if and only if $\alpha X+\beta \leqslant_r^{\mathrm{R}} \alpha Y+\beta$ for all $\alpha \geqslant 0$ and $\beta \geqslant 0$.

We briefly show (iii). Note that $\leqslant_r^{\mathrm{R}}$ is invariant to positive scaling. It suffices to show (iii) for $\alpha=1$. The $\Leftarrow$ implication is trivial. For the $\Rightarrow$ implication, note that the function $x \mapsto (x^r+\beta)^{1/r}$ is concave on $[0,\infty)$. Using Corollary 1, we have $X^{1/r} \leqslant_{\mathrm{SSD}} Y^{1/r}$, which leads to $(X+\beta)^{1/r} \leqslant_{\mathrm{SSD}} (Y+\beta)^{1/r}$ because of the concavity of $x \mapsto (x^r+\beta)^{1/r}$. Using Corollary 1 again, we obtain $X+\beta \leqslant_r^{\mathrm{R}} Y+\beta$.

Below we present a slightly stronger result on the translation property of $(1+r)_{\mathrm{R}}$-SD than (iii) above, allowing for a possibly negative location shift. Proposition 4 will be useful in the proof of Theorem 3.

Proposition 4. *Let X and Y be two random variables taking values on $[a,b]$. Then $X \leqslant_r^{\mathrm{R}} Y$ if and only if $X+\beta \leqslant_{s(\beta)}^{\mathrm{R}} Y+\beta$ for all $\beta \geqslant -a$ where*

$$s(\beta)=\max\left\{\frac{ra}{a+\beta-r\beta},\ \frac{rb}{b+\beta-r\beta}\right\} \leqslant 1.$$

A preference $\precsim$ is *scale invariant* if for all X, Y, $X \precsim Y$ implies $\alpha X \precsim \alpha Y$ for all $\alpha>0$. A preference $\precsim$ is *translation invariant* if for all X, Y, $X \precsim Y$ implies $X+\beta \precsim Y+\beta$ for all $\beta \in \mathbb{R}$. The preference $\precsim$ is *lower semi-continuous* if $|X_n-X| \to 0$ uniformly and $X_n \precsim Y$ for each n implies $X \precsim Y$ and it is *upper semi-continuous* if $|Y_n-Y| \to 0$ uniformly and $X \precsim Y_n$ for each n implies $X \precsim Y$.

Theorem 3. *Let $\precsim$ be a lower or upper semi-continuous preference on the set of bounded random variables, denoted by $\mathcal{X}$.*

(i) Suppose that the preference $\precsim$ is scale invariant. For each $c \in (0,1]$, $\precsim$ is monotone in $\leqslant_c^{\mathrm{A}}$ if and only if it is monotone in SSD.

(ii) Suppose that the preference $\precsim$ is translation invariant. For each $r \in (0,1]$, $\precsim$ is monotone in $\leqslant_r^{\mathrm{R}}$ if and only if it is monotone in SSD.

Theorem 3 implies, in particular, that monotonicity in $(1+c)_{\mathrm{A}}$-SD (or $(1+c)_{\mathrm{R}}$-SD) is equivalent to monotonicity in SSD for any preference on $\mathcal{X}$ modeled by Yaari's dual utility, which is scale/translation invariant and lower/upper semi-continuous. Note that translation invariance is not compatible with CPT or RDU unless the associated utility function is linear. We have seen this equivalence in Section 3. The next result concerns a comparison of $(1+c)_{\mathrm{A}}$-SD and $(1+\gamma)_{\mathrm{S}}$-SD for prospects that do not have a finite expected loss (these prospects cannot be compared by $(1+r)_{\mathrm{R}}$-SD since they are not non-negative).

Proposition 5. *For any random variable X with $\mathbb{E}[X_-] = \infty$,*

(i) $X \leqslant_\gamma^{\mathrm{S}} Y$ for all $\gamma \in (0,1]$ and Y.

(ii) For $0 \leqslant c' < c < 1$, there exists Y such that $X \leqslant_c^{\mathrm{A}} Y$ and $X \not\leqslant_{c'}^{\mathrm{A}} Y$.

Proposition 5 shows that $(1+c)_{\mathrm{A}}$-SD is able to distinguish between prospects with infinite expected losses, whereas $(1+\gamma)_{\mathrm{S}}$-SD is completely blind in such situations, just like SSD. In contrast, any heavy-tailed random variable X and any constant $d \in \mathbb{R}$ cannot be compared by $(1+c)_{\mathrm{A}}$-SD unless $d \leqslant X$, which is the case of FSD.

Proposition 6. *For any random variable X and $d \in \mathbb{R}$, if $\mathbb{E}[e^{\lambda X}] = \infty$ for some $\lambda > 0$, then X and d do not dominate each other in $(1+c)_{\mathrm{A}}$-SD for $c \leqslant 1/(1+\lambda)$ unless $d \leqslant X$.*

In what follows, for notational convenience, we will write fractional SD between distribution functions F and G, which should be understood as the fractional SD between the corresponding random variables with these distributions.

Example 4. We will see that, in contrast to $(1+\gamma)_{\mathrm{S}}$-SD, $(1+c)_{\mathrm{A}}$-SD cannot distinguish log-normal distributions or Pareto distributions except for the case of FSD. Technical details of the statements here are given in Section A.3.

(1) (Log-normal distribution.) The log-normal distribution is widely used in both finance and income/wealth modeling. Let $X_k = \exp(\mu_k + \sigma_k Z)$, $k = 1, 2$, where Z is a standard normal distributed prospect with distribution Φ, $\sigma_1 \geqslant \sigma_2$, and $\mu_2 - \mu_1 \geqslant (\sigma_1^2 - \sigma_2^2)/2$, which guarantee

that $X_1 \leqslant_{\mathrm{SSD}} X_2$; see e.g., Theorem 5 of Levy (1973). In the case $\sigma_1 = \sigma_2$, either $X_1 \leqslant_{\mathrm{FSD}} X_2$ or $X_2 \leqslant_{\mathrm{FSD}} X_1$. If $\sigma_1 > \sigma_2$, then $X_1 \leqslant^{\mathrm{S}}_{\gamma} X_2$ for $\gamma \geqslant \gamma_{\min}$ where

$$\gamma_{\min} = \frac{\int_{t_0}^{\infty} e^x \left(\Phi\left(\frac{x-\mu_2}{\sigma_2}\right) - \Phi\left(\frac{x-\mu_1}{\sigma_1}\right)\right) \mathrm{d}x}{\int_{-\infty}^{t_0} e^x \left(\Phi\left(\frac{x-\mu_1}{\sigma_1}\right) - \Phi\left(\frac{x-\mu_2}{\sigma_2}\right)\right) \mathrm{d}x} \quad \text{with } t_0 = \frac{\mu_1\sigma_2 - \mu_2\sigma_1}{\sigma_2 - \sigma_1};$$

moreover, $X_1 \leqslant^{\mathrm{R}}_{r} X_2$ if and only if $\sigma_1 \geqslant \sigma_2$, and $r(\mu_2 - \mu_1) \geqslant (\sigma_1^2 - \sigma_2^2)/2$. In contrast, $X_1 \not\leqslant^{\mathrm{A}}_{c} X_2$ for any $c < 1$ in this case.

(2) (Pareto/Lomax distribution.) Let F_k be a distribution given by

$$F_k(x) = 1 - \left(1 + \frac{x}{\mu_k}\right)^{-\alpha_k}, \quad x \geqslant 0, \quad k = 1, 2,$$

where $\mu_k > 0$, $k = 1, 2$, $\alpha_1 \geqslant \alpha_2 > 0$. If $\alpha_1 = \alpha_2$, then either $F_1 \leqslant_{\mathrm{FSD}} F_2$ or $F_2 \leqslant_{\mathrm{FSD}} F_1$ depending on $\mu_1 \leqslant \mu_2$ or $\mu_1 \geqslant \mu_2$. Next, we assume $\alpha_1 > \alpha_2$ and consider the following cases.

(i) If $\mu_1 \leqslant \mu_2$ or $\mu_1 > \mu_2$, $\alpha_1\mu_2 - \alpha_2\mu_1 \geqslant 0$, then $F_1 \leqslant_{\mathrm{FSD}} F_2$.

(ii) $\mu_1 > \mu_2$, $\alpha_1\mu_2 - \alpha_2\mu_1 < 0$.

(ii.a) If $1 \geqslant \alpha_2$ or $\alpha_2 > 1$, $\mu_1/(\alpha_1 - 1) < \mu_2/(\alpha_2 - 1)$, then $F_1 \not\leqslant_{\mathrm{SSD}} F_2$ and $F_2 \not\leqslant_{\mathrm{SSD}} F_1$.

(ii.b) If $\alpha_2 > 1$, $\mu_1/(\alpha_1 - 1) \geqslant \mu_2/(\alpha_2 - 1)$, then $F_2 \leqslant^{\mathrm{S}}_{\gamma} F_1$ for $\gamma \geqslant \gamma_{\min}$ where

$$\gamma_{\min} = \frac{\int_{x_0}^{\infty} \frac{\mu_2^{\alpha_2}}{(x+\mu_2)^{\alpha_2}} - \frac{\mu_1^{\alpha_1}}{(x+\mu_1)^{\alpha_1}} \,\mathrm{d}x}{\int_{0}^{x_0} \frac{\mu_1^{\alpha_1}}{(x+\mu_1)^{\alpha_1}} - \frac{\mu_2^{\alpha_2}}{(x+\mu_2)^{\alpha_2}} \,\mathrm{d}x}.$$

We also have $F_2 \leqslant^{\mathrm{R}}_{r} F_1$ for $r \geqslant r_{\min}$ where $r_{\min}$ is the solution to the equation

$$\int_0^{\infty} \left(1 + \frac{x^r}{\mu_1}\right)^{-\alpha_1} - \left(1 + \frac{x^r}{\mu_2}\right)^{-\alpha_2} \mathrm{d}x = 0.$$

In contrast, for any $\alpha_1 > \alpha_2 > 0$ and $c \in (0, 1)$, $F_2 \not\leqslant^{\mathrm{A}}_{c} F_1$.

5 Estimation of the parameter in a fractional SD relationship

In this and the next sections, we study the estimations and real-data applications of the indexes γ in $(1+\gamma)_{\mathrm{S}}$-SD and c in $(1+c)_{\mathrm{A}}$-SD. The index r in $(1+r)_{\mathrm{R}}$-SD shares some similarity with that of c and is omitted. Since both relations $(1+c)_{\mathrm{A}}$-SD and $(1+\gamma)_{\mathrm{S}}$-SD become stronger as c and γ decrease, it is natural to identify the smallest values (or infima) of γ and c such that $F \leqslant^{\mathrm{S}}_{\gamma} G$ and $F \leqslant^{\mathrm{A}}_{c} G$ hold for two given distributions F and G. We denote these two numbers by $\gamma_{\min}$ and $c_{\min}$, respectively, assuming that they exist (if they do not, then we use the infima). In practice, we often need to determine estimates of $\gamma_{\min}$ and $c_{\min}$ based on observations $X_1, \dots, X_n \sim F$ and $Y_1, \dots, Y_m \sim G$. Denote by F_n and G_m the empirical cumulative distributions (cdfs) of $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$, respectively. We assume that $F_n \to F$ and $G_m \to G$ as $n, m \to \infty$ in probability. Clearly, this is the minimum requirement of any meaning approximation of the true distributions using the empirical ones, and it is satisfied by, for instance, iid or α-mixing stationary data.

We first give an equivalent characterization of $(1+c)_{\mathrm{A}}$-SD. We say that two distributions F and G are single-crossing at $x_0 \in \mathbb{R}$ if either $F - G \leqslant 0$ on $(-\infty, x_0)$ and $F - G \geqslant 0$ on (x_0, ∞), or $F - G \geqslant 0$ on $(-\infty, x_0)$ and $F - G \leqslant 0$ on (x_0, ∞). If, in addition, $F \leqslant_{\mathrm{SSD}} G$, then only the latter case is possible. For a random variable X, define

$$X_p = F^{-1}(pU), \quad p \in [0,1],$$

where F is the distribution function of X. If p is small, the random variable X_p can be interpreted as the tail risk of X; see Liu and Wang (2021).

Proposition 7. *For random variables X and Y with respective distributions F and G and finite exponential moments, $X \leqslant_c^{\mathrm{A}} Y$ if and only if $c \geqslant c_{\min} = 1/(1+\lambda_{\max})$, where*

$$\lambda_{\max} = \sup\{\lambda \geqslant 0 : g(\lambda) \geqslant 0\}, \quad g(\lambda) = \inf_{p \in (0,1)} \left(\mathbb{E}[e^{\lambda Y_p}] - \mathbb{E}[e^{\lambda X_p}] \right). \tag{13}$$

In particular, if F and G are single-crossing at some point, then $\lambda_{\max}$ is the unique solution $\lambda > 0$ of the equation $\mathbb{E}[e^{\lambda X}] = \mathbb{E}[e^{\lambda Y}]$.

By Proposition 7, we propose an estimator of $c_{\min}$ as follows,

$$\hat{c}_{\min} = (1 + \hat{\lambda}_{\max})^{-1} \quad \text{and} \quad \hat{\lambda}_{\max} = \sup\{\lambda \geqslant 0 : \hat{g}(\lambda) \geqslant 0\},$$

where

$$\hat{g}(\lambda) = \inf_{p \in (0,1)} \left(\mathbb{E}_{Y \sim G_m}[e^{\lambda Y_p}] - \mathbb{E}_{X \sim F_n}[e^{\lambda X_p}] \right).$$

For a given dataset, computation and properties of $\hat{c}_{\min}$ may not be easy to analyze in general since two layers of optimization are involved when finding the largest $\lambda > 0$ such that $e^{\lambda Z} \leqslant_{\mathrm{SSD}} e^{\lambda W}$ holds for $Z \sim F_n$ and $W \sim G_m$. We are not aware of a simple and efficient procedure for accomplishing this task for general underlying distributions. Technically, due to the possibly complicated relationship between F and G, it is not obvious how to establish consistency of the estimator $\hat{c}_{\min}$. For the above reason, we will investigate the simpler case of single-crossing, a popular setup for distributions satisfying an SSD relationship.

Suppose that F and G are single-crossing. We can set $\hat{c}_{\min} = (1 + \hat{\lambda}_{\max})^{-1}$, where $\hat{\lambda}_{\max}$ is the largest (usually unique) λ satisfying $\mathbb{E}^{F_n}[e^{\lambda X}] = \mathbb{E}^{G_m}[e^{\lambda Y}]$, and we set $\hat{\lambda}_{\max} = 0$ if there is no solution to the above equation. Under suitable regularity conditions, the probability that F_n and G_m are single-crossing tends to 1 as $n, m \to \infty$.

Below, we say that F and G are strictly single-crossing if they are single-crossing, and there exists $t_0 \in (0, 1)$ such that for any $\varepsilon > 0$, there exists $\delta > 0$ such that $|F^{-1}(t) - G^{-1}(t)| > \varepsilon$ for all x with $|t - t_0| > \delta$. That is, the curves of F and G cross at one point, and they depart before and after that point.

Proposition 8. *Let F and G be two strictly single-crossing and compactly supported continuous distributions. Then $\hat{c}_{\min}$ is a consistent estimator of $c_{\min}$; that is, $\hat{c}_{\min} \to c_{\min}$ in probability as $n, m \to \infty$.*

Estimating $\gamma_{\min}$ is simpler than estimating $c_{\min}$. We follow the equivalent characterization given by Theorem 2.4 of Müller et al. (2017). Given two distributions F and G, $F \leqslant_{\gamma}^{\mathrm{S}} G$ if and only if the following statement holds

$$\int_{-\infty}^{t} (G(x) - F(x))_+ \mathrm{d}x \leqslant \gamma \int_{-\infty}^{t} (F(x) - G(x))_+ \mathrm{d}x \quad \text{for all } t \in \mathbb{R}.$$

This implies that[4]

$$\gamma_{\min} = \sup_{t \in \mathbb{R}} \frac{\int_{-\infty}^{t} (G(x) - F(x))_+ \mathrm{d}x}{\int_{-\infty}^{t} (F(x) - G(x))_+ \mathrm{d}x}.$$

Based on this observation, we propose the following $\hat{\gamma}_{\min}$ as one estimation of $\gamma_{\min}$:

$$\hat{\gamma}_{\min} = \min\{\tilde{\gamma}_{\min}, 1\}, \quad \tilde{\gamma}_{\min} = \max_{t \in \mathbb{R}} \frac{\int_0^t (G_m(x) - F_n(x))_+ \, \mathrm{d}x}{\int_0^t (F_n(x) - G_m(x))_+ \, \mathrm{d}x}.$$

In contrast to $\hat{c}_{\min}$, the consistency of $\hat{\gamma}_{\min}$ is easy to establish as it involves only one layer of optimization.

Proposition 9. *Let F and G be two compactly supported continuous distributions, and let ℓ_F and ℓ_G be the left endpoints of the supports of F and G, respectively. If $\ell_F < \ell_G$, then $\hat{\gamma}_{\min}$ is a consistent estimator of $\gamma_{\min}$, that is, $\hat{\gamma}_{\min} \to \gamma_{\min}$ in probability as $n, m \to \infty$.*

Simulations of $\hat{\gamma}_{\min}$ and $\hat{c}_{\min}$ for normal distributions are presented in Appendix B. Although normal distributions do not satisfy the condition of compact support in Propositions 8 and 9, they satisfy the single-crossing property, and $\gamma_{\min}$ and $c_{\min}$ have explicit formulas. The simulation results confirm that estimators $\hat{\gamma}_{\min}$ and $\hat{c}_{\min}$ work quite well in this setting.

[4] A similar index called the measure for partial stochastic dominance $\alpha(F, G)$ is defined by Eq. (10) in Kamihigashi and Stachurski (2020).

6 Real data example

We compare the log-return and the daily change[5] in the S&P500 index, the Dow Jones Industrial (DJI), and NASDAQ of 2005 (251 data points) with those of 2008 (252 data points). For each index, we denote the empirical cdfs of the 2005 and 2008 samples by F_{05} and F_{08}, respectively. As shown in Figure 1, for each index, F_{05} and F_{08} are single-crossing and F_{05} has a larger mean than F_{08}, which implies that $F_{08} \not\leqslant_{\rm FSD} F_{05}$ and $F_{08} \leqslant_{\rm SSD} F_{05}$[6]. Similarly, we also compare the log-return and daily change of S&P500, DJI and NASDAQ of 2019 (Q1-Q2, 123 data points from January 3, 2019 to June 28, 2019) and 2020 (Q1-Q2, 124 data points from January 3, 2020 to June 30, 2020). For each index, we denote these distributions by F_{19} and F_{20}, respectively. In Figure 2, for each index, F_{19} and F_{20} are single-crossing and F_{19} has a larger mean than F_{20}, which implies that $F_{20} \not\leqslant_{\rm FSD} F_{19}$ and $F_{20} \leqslant_{\rm SSD} F_{19}$. We choose the pairs 2005-2008 and 2019-2020 because the preference of 2005 over 2008 and 2019 over 2020 is arguably intuitive, and they satisfy the single-crossing property and a mean inequality, which implies an SSD order. In Table 1, we report the estimated values of $\gamma_{\min}$ and $c_{\min}$ for the 2005-2008 data and 2019-2020 data using the methods outlined in Section 5.

From the results on $(1+\gamma)_{\rm S}$-SD, we observe that, either in log-returns or in absolute value, compared to 2008, the log-return distributions in 2005 are preferred by all investors who are monotone in $(1+0.75)_{\rm S}$-SD, and compared to 2020, the log-return distributions in 2019 are preferred by all investors who are monotone in $(1+0.9)_{\rm S}$-SD. Between the two pairs 2005-2008 and 2019-2020 (Q1-Q2), the strength of dominance is comparable for S&P500 and DJI, and the dominance in 2005-2008 is stronger than 2019-2020 for NASDAQ.

From results on $(1+c)_{\rm A}$-SD, the results are either very close to 0 (using log-return) or very close to 1 (using daily change). Since the estimates of $c_{\min}$ are sensitive to scaling, it is not clear to us how conclusions can be drawn from these numbers.

7 An application of portfolio selection

We present an application of Theorem 1 in a portfolio selection problem, similar to the setting considered by Chew et al. (1987). For simplicity, we consider an investor with a DT(h) preference, and we denote the numerical representation of DT(h) by D_h, i.e., $D_h(X) = \int x \, \mathrm{d}h(F(x))$ where F

[5]The daily change is the difference between the indices on two consecutive trading days. These data are not iid, but the methods in Section 5 do not require an iid assumption.

[6]For two cdfs F and G, the dominance between F and G means the dominance between X and Y, where X and Y are two random variables having cdfs F and G, respectively.

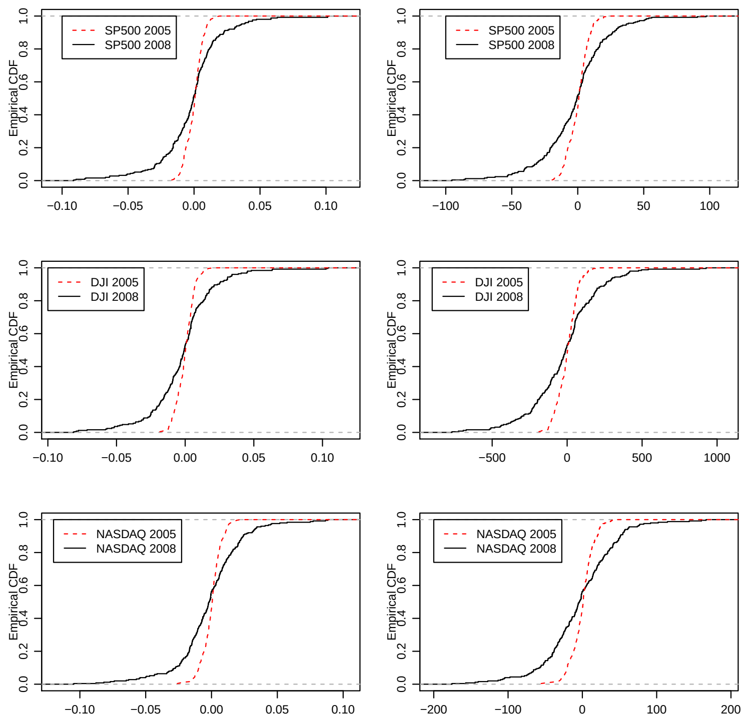

Figure 1: Empirical cdfs of 2005 versus 2008. Left panel: log-return; right panel: daily change.

is the distribution of X. Suppose that the investor has an optimal portfolio problem

$$\alpha^*(X) = \underset{\alpha \in [0,\alpha_0]}{\arg\max}\, D_h(w + \alpha X - c(\alpha))$$

where $w \in [0,\infty)$ is the initial wealth of the investor, X is the future price of an asset, $\alpha_0 \in (0,\infty)$, and $c : [0,\infty) \to [0,\infty)$ is a cost function. The value $c(\alpha)$ of the cost function represents the price paid to purchase α units of the asset. We assume that c is increasing and strictly convex; an example is $c(x) = ax^2$ for some $a > 0$. These assumptions are consistent with the intuition that the marginal cost is increasing due to transaction fees or limited liquidity (e.g., Föllmer and Schied (2002)). In what follows, two assets are normalized so that they share the same cost function.

Proposition 10. *If $\gamma \in (0,1)$, $P_h \geqslant \gamma$ where P_h is given by (6), and $X_1 \leqslant^{\mathrm{S}}_{\gamma} X_2$, then $\alpha^*(X_1) \leqslant \alpha^*(X_2)$.*

To interpret the result in Proposition 10, we take the example in Section 6 by assuming that X_1 and X_2 represent the return of 1 unit of the S&P 500 index using the distributions F_{08} and F_{05}, respectively. As we have seen in Section 6, we have $X_1 \leqslant^{\mathrm{S}}_{0.7} X_2$. If a DT($h$) decision maker satisfies $P_h \geqslant 0.7$ (see examples of such h in Example 2), then she would invest more in the index if the market follows the empirical distribution in 2005 and invest less if the market follows the empirical

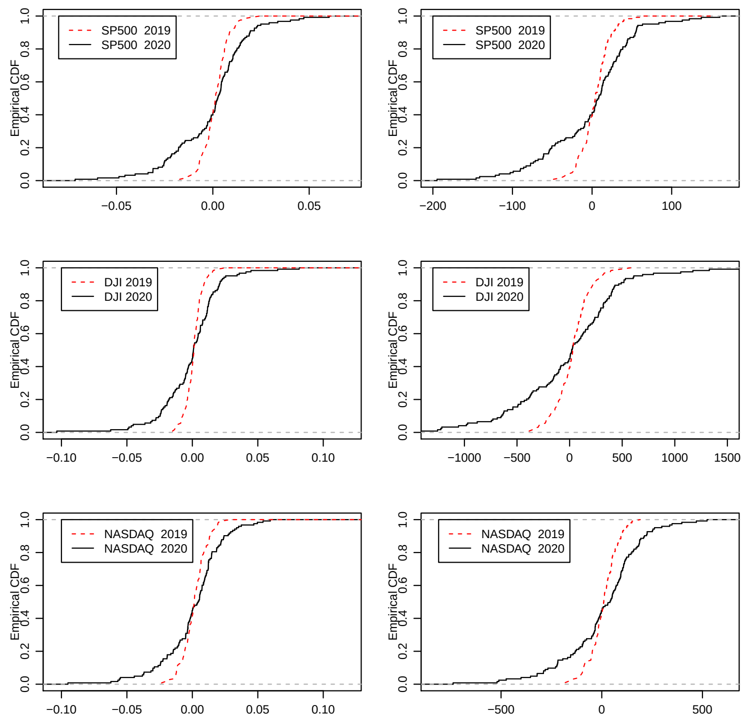

Figure 2: Empirical cdfs of 2019 versus 2020, Q1-Q2. Left panel: log-return; right panel: daily change.

distribution in 2008. This conclusion holds true for some risk-seeking investors, as $P_h \geqslant 0.7$ allows for some risk-seeking DT models, i.e., those with a convex h.

8 Concluding remarks

As we have seen from Corollary 1, the concepts of $(1+c)_{\mathrm{A}}$-SD and $(1+r)_{\mathrm{R}}$-SD can be roughly seen as a logarithmic version and a power version of SSD, respectively, which have a clear connection to the classic framework of risk aversion, in particular, to coefficients of risk aversion. On the other hand, $(1+\gamma)_{\mathrm{S}}$-SD offers a significantly different technical framework than the classic ones, thus allowing for applications in more situations of behavioral decision analysis. Both $(1+c)_{\mathrm{A}}$-SD and $(1+r)_{\mathrm{R}}$-SD are included in the class of v-SD, which allows for construction of more general notions of fractional SD.

If behavioral decision models involving (subjective) probability distortion such as RDU or CPT are used, then $(1+\gamma)_{\mathrm{S}}$-SD is more suitable than v-SD, as it correctly reflects the role of the probability distortion in the comparison of risks. On the other hand, $(1+c)_{\mathrm{A}}$-SD and $(1+r)_{\mathrm{R}}$-SD are, in the sense of Theorem 1, blind to probability distortion.

One could nevertheless construct fractional SD directly from RDU (or DT, CPT) preferences.

		2005 vs 2008		2019 vs 2020 (Q1-Q2)	
		$\gamma_{\min}$	$c_{\min}$	$\gamma_{\min}$	$c_{\min}$
S&P500	log-return	0.6871	0.1184	0.6593	0.0824
	daily change	0.6688	0.9918	0.6706	0.9961
DJI	log-return	0.7424	0.1380	0.6806	0.1000
	daily change	0.7348	0.9999	0.6582	0.9997
NASDAQ	log-return	0.7077	0.1206	0.8309	0.1980
	daily change	0.6992	0.9958	0.8958	0.9997

Table 1: Estimated values of $\gamma_{\min}$ and $c_{\min}$.

For instance, analogously to (2), let

$$\mathscr{R}_{v,g} = \{(u,h) \mid u = w \circ v,\ h = f \circ g \text{ for some increasing concave } w \text{ and } f\},$$

and we can define (v,g)-SD by $X \leqslant_{v,g} Y \Leftrightarrow V_{u,h}(X) \leqslant V_{u,h}(Y)$ for all $(u,h) \in \mathscr{R}_{v,g}$. With this formulation, the roles of the utility function and the probability perception function become symmetric. It is a bit surprising that for $(1+\gamma)_{\mathrm{S}}$-SD, the roles of the utility function and the probability perception function in RDU are indeed symmetric, although the formulation of $(1+\gamma)_{\mathrm{S}}$-SD only involves utility functions; the same holds true for the SSD conditions of RDU.

Our paper is the first to connect the important behavioral models of RDU and CPT in decision theory and the recent notions of fractional stochastic dominance. As a potential benefit, our results allow for an economist to know what kind of decision models to use when fractional SD is assumed or empirically observed. Using these relations, one can pin down optimal choices in some applications without specifying the preference model. To present a concise analysis in this paper, we did not include other types of SD, such as the prospect SD of Baucells and Heukamp (2006) and the continuum of SD of Fishburn (1976, 1980). These generalizations can be studied in the future. Fractional SD is also studied in other decision models by Yang et al. (2022).

Some other comparisons between the three notions of fractional SD are drawn from our further analyses in Sections 4-6. If the risk comparison is naturally scalable, such as monetary amounts possibly via exchange rates, then $(1+\gamma)_{\mathrm{S}}$-SD and $(1+r)_{\mathrm{R}}$-SD are better to use (Theorem 3). On the other hand, if it is desirable that the risk comparison is invariant to location shift, then $(1+\gamma)_{\mathrm{S}}$-SD and $(1+c)_{\mathrm{A}}$-SD are more suitable. If the risks to compare do not have a finite expected loss (statistically, this means the situation that the first moment is difficult to estimate, even if it exists), then $(1+c)_{\mathrm{A}}$-SD is useful whereas $(1+\gamma)_{\mathrm{S}}$-SD is not (Proposition 5). For prospects that take both positive and negative values, $(1+r)_{\mathrm{R}}$-SD is not properly defined. For distributions of prospects with no exponential moments such as log-normal or Pareto distributions, $(1+\gamma)_{\mathrm{S}}$-SD and

$(1 + r)_{\mathrm{R}}$-SD are more useful than $(1 + c)_{\mathrm{A}}$-SD (Proposition 6 and Example 4). When it comes to estimation of the parameters c and γ from real data, $(1 + \gamma)_{\mathrm{S}}$-SD is often more convenient to work with, and its estimates are easier to interpret, at least in our examples in Sections 5 and 6.

References

Arrow, K.J. (1974). *Essays in the Theory of Risk-bearing.* North-Holland, New York.

Baucells, M. and Heukamp, F. H. (2006). Stochastic dominance and cumulative prospect theory. *Management Science*, **52**(9), 1409–1423.

Chateauneuf, A., Cohen, M. and Meilijson, I. (2005). More pessimism than greediness: a characterization of monotone risk aversion in the rank-dependent expected utility model. *Economic Theory*, **25**, 649–667.

Chew, S. H., Karni, E. and Safra, Z. (1987). Risk aversion in the theory of expected utility with rank dependent probabilities. *Journal of Economic Theory*, **42**, 370–381.

Fishburn, P. C. (1976). Continua of stochastic dominance relations for bounded probability distributions. *Journal of Mathematical Economics*, **3**, 295–311.

Fishburn, P. C. (1980). Continua of stochastic dominance relations for unbounded probability distributions. *Journal of Mathematical Economics*, **7**, 271–285.

Föllmer, H. and Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and Stochastics*, **6**(4), 429–447.

Hadar, J. and Russell, W. (1969). Rules for ordering uncertain prospects. *American Economic Review.* **59**(1), 25–34.

Hanoch, G. and Levy, H. (1969). The efficiency analysis of choices involving risk. *Review of Economic Studies.* **36**(3), 335–346.

Harrison, G. W. (1986). An experimental test for risk aversion. *Economics Letters*, **21**(1), 7–11.

Huang, R. J., Tzeng, L. Y. and Zhao, L. (2020). Fractional degree stochastic dominance. *Management Science*, **66**(10), 4630–4647.

Levy, H. (1973). Stochastic dominance among log-normal prospects. *International Economic Review*, **14**(3), 601–614.

Liu, F. and Wang, R. (2021). A theory for measures of tail risk. *Mathematics of Operations Research*, **46**(3), 1109–1128.

Liu, P. and Schied, A. and Wang, R. (2021). Distributional transforms, probability distortions, and their applications. *Mathematics of Operations Research*, **46**(4), 1490-1512.

Kamihigashi, T. and Stachurski, J. (2020). Partial stochastic dominance via optimal transport. *Operations Research Letters*, **48**(5), 584–586.

Kimball, M. S. (1993). Standard risk aversion. *Econometrica*, **61**, 589–611.

Meyer, J. (1977). Second degree stochastic dominance with respect to a function. *International Economic Review*, **18**(2) 477–487.

Müller, A., Scarsini, M., Tsetlin, I. and Winkler. R. L. (2017). Between first and second-order stochastic dominance. *Management Science*, **63**(9), 2933–2947.

Müller, A., Scarsini, M., Tsetlin, I. and Winkler. R. L. (2021). Technical note—Ranking distributions when only means and variances are known. *Operations Research*, published online.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, **32**, 122–136.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization*, **3**, 323–343.

Rabin, M. (2000). Risk aversion and expected-utility theory: A calibration theorem. *Econometrica*, **68**(5), 1281–1292.

Rabin, M. and Thaler, R. H. (2001). Anomalies risk aversion. *Journal of Economic Perspetives*, **15**(1), 219–232.

Rothschild, M. and Stiglitz, J. E. (1970). Increasing risk: I. A definition. *Journal of Economic Theory*, **2**(3), 225–243.

Schmidt, U. and Traub, S. (2002). An experimental test of loss aversion. *Journal of Risk and Uncertainty*, **25**(3), 233–249.

Schmidt, U. and Zank, H. (2008). Risk aversion in cumulative prospect theory. *Management Science*, **54**, 208–216.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. Springer Series in Statistics.

Strassen, V. (1965). The existence of probability measures with given marginals. *Annals of Mathematical Statistics*, **36**(2), 423–439.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, **5**(4), 297–323.

Wang, R. and Wu, Q. (2020). Dependence and risk attitudes: An equivalence. *SSRN*: 3707709.

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, **55**, 95–115.

Yang, J., Zhao, C., Chen, W., Zhou, D., and Han, S. (2022). Fraction-degree reference dependent stochastic dominance. *Methodology and Computing in Applied Probability*, **24**, 1193–1219.

A Proofs of all results

A.1 Proof of Propositions 2 and 3 in Section 2

We first present the proof of Proposition 2, which says that unlike the other notions of fractional SD considered in this paper, $(1+\gamma)_{\mathrm{S}}$-SD of Müller et al. (2017) cannot be formulated via SSD.

Proof of Proposition 2. We show the result by contradiction. Denote by $\mathcal{M}_{[a,b]}$ the set of all cdfs with support in $[a,b]$ and F and G the cdfs of X and Y, respectively. For $\gamma < 1$, suppose there exists a $v : \mathbb{R} \to \mathbb{R}$ and $h : [0,1] \to [0,1]$ such that (3) holds. Let $T_\gamma : \mathcal{M}_{[a,b]} \to \mathcal{M}$ be $T_\gamma(F) = h \circ F \circ v$, where $\mathcal{M}$ is the set of distributions. Since both $\leqslant_\gamma^{\mathrm{S}}$ and $\leqslant_{\mathrm{SSD}}$ are reflexive, we have that the transform T_γ is a one-to-one transform, that is, h and v are two one-to-one functions. First consider the distribution δ_x representing for the point-mass at $x \in \mathbb{R}$, we have $T_\gamma(\delta_x) = h(\delta_x \circ v) = h(\mathbb{1}_{\{v(\cdot) \geqslant x\}}) = h(0)\mathbb{1}_{\{v(\cdot)<x\}} + h(1)\mathbb{1}_{\{v(\cdot)\geqslant x\}} \in \mathcal{M}$. Hence, we have $\{h(0), h(1)\} = \{0, 1\}$. Comparing δ_x and δ_y, $a \leqslant x < y \leqslant b$, by (3), we have that either one of the following two cases holds: (1) $v(x) < v(y)$, $h(0) = 0$ and $h(1) = 1$; (2) $v(x) > v(y)$, $h(0) = 1$ and $h(1) = 0$. Without loss of generality (wlog), assume that case (1) holds, that is, v is strictly increasing, $h(0) = 0$ and $h(1) = 1$. Define $u : [a,b] \to \mathbb{R}$ as $u(x) = v^{-1}(x) = \inf\{y : v(y) \geqslant x\}$. Then by that the transform T_γ is a one-to-one transform, and $T_\gamma(\delta_x)$ is the point mass at $u(x)$, we have u is strictly increasing. Hence, we have v is continuous and thus, v and u are both increasing and continuous.

Consider the distribution $F_p = p\delta_0 + (1-p)\delta_1$, $p \in [0,1]$. Note that F_p is decreasing in p with respect to SSD and thus, we have $T_\gamma(F) = h(p)\delta_{u(0)} + (1-h(p))\delta_{u(1)}$ is decreasing in p with respect to $(1+\gamma)_{\mathrm{S}}$-SD. Hence, h is also strictly increasing. For $p \in (0,1)$, define $G_n = p_n\delta_{1/n} + (1-p_n)\delta_1$, where $p_n = p + \gamma p/(n-1) \in (p, 1)$ for n large enough. Then $F_p \leqslant_\gamma G_n$ each n such that $p_n \in (p,1)$. By (3), we have $T_\gamma(F_p) \leqslant_{\mathrm{SSD}} T_\gamma(G_n)$, that is, $h(p)\delta_{u(0)} + (1-h(p))\delta_{u(1)} \leqslant_{\mathrm{SSD}} h(p_n)\delta_{u(1/n)} + (1-h(p_n))\delta_{u(1)}$. Note that u is continuous at 0 and $h(p_n)\delta_{u(1/n)} + (1-h(p_n))\delta_{u(1)}$ converges to $h(p+)\delta_0 + (1-h(p+))\delta_{u(1)}$ in distribution and the expectation also converges as $n \to \infty$, where $h(p+) = \lim_{q\downarrow p} h(p)$. Since $\leqslant_{\mathrm{SSD}}$ is continuous with respect to the above convergence, we conclude that $h(p)\delta_{u(0)} + (1-h(p))\delta_{u(1)} \leqslant_{\mathrm{SSD}} h(p+)\delta_{u(0)} + (1-h(p+))\delta_{u(1)}$. This implies that $h(p+) \leqslant h(p)$. Similarly, we can show $h(p-) = \lim_{q\uparrow p} h(p) \geqslant h(p)$. Hence, we have h is continuous on $(0,1)$. We also can similarly show h is continuous on $[0,1]$.

Define $g = h^{-1}$ and $U_\gamma(F) = T_\gamma^{-1}(F) = g \circ F \circ u$. Then g is increasing and continuous in $[0,1]$ with $g(0) = 0$, $g(1) = 1$, and U_γ is a transform from $\mathcal{M}$ to $\mathcal{M}_{[a,b]}$. For $x \in [-2,2]$, let F_x be the distribution of a random variable X_x which is uniformly distributed on $\{x-1, x+1\}$. Obviously,

$F_x \leqslant_{\mathrm{SSD}} \delta_x$, and thus, $U_\gamma(F_x) \leqslant_\gamma U_\gamma(\delta_x)$. Hence, for $x \in \mathbb{R}$,

$$\Delta_u(x) g_{1/2} \leqslant \gamma \quad \text{with} \quad \Delta_u(x) = \frac{u(x+1) - u(x)}{u(x) - u(x-1)}, \quad g_{1/2} = \frac{1 - g(1/2)}{g(1/2)}.$$

If $\Delta_u(x) g_{1/2} < \gamma$, then there exists $x_0 < x$ such that $\Delta_u(x_0) g_{1/2} = \gamma$. This implies that $U_\gamma(F_x) \leqslant_\gamma U_\gamma(\delta_{x_0})$. Then by (3), we have $F_x \leqslant_{\mathrm{SSD}} \delta_{x_0}$ which yields a contradiction with that $\mathbb{E}[X_x] = x > x_0$. Therefore, we have $\Delta_u(x) g_{1/2} = \gamma$. Similarly, compare the distributions $(F_x + G)/2$ and $(\delta_x + G)/2$ with $G = \delta_{-4}, \delta_4$ and $(\delta_{-4} + \delta_4)/2$, respectively, we have

$$\Delta_u(x) \cdot \frac{g(1/2) - g(1/4)}{g(1/4)} = \Delta_u(x) \cdot \frac{g(3/4) - g(1/2)}{g(1/2) - g(1/4)} = \Delta_u(x) \cdot \frac{g(1) - g(3/4)}{g(3/4) - g(1/2)} = \gamma.$$

Those imply $\frac{g(1/2)-g(1/4)}{g(1/4)} = \frac{g(3/4)-g(1/2)}{g(1/2)-g(1/4)} = \frac{g(1)-g(3/4)}{g(3/4)-g(1/2)} = \frac{1-g(1/2)}{g(1/2)}$ and thus, $g(x) = x$ for $x = 0, 1/4, 1/2, 3/4, 1$. Similarly, we can show $g(k/2^n) = k/2^n$, $k = 1, \ldots, 2^n$, $n \in \mathbb{N}$. By monotonicity, we have $g(x) = x$ for $x \in [0,1]$, and thus, $h(x) = x$ for $x \in [0,1]$. This then implies $\Delta_u(x) = \gamma$ for $x \in [-2,2]$. Define Z such that $\mathbb{P}(Z = x-1) = 1 - \mathbb{P}(Z = x+2) = 2/3$. Then we have $Z \leqslant_{\mathrm{SSD}} x$ which is equivalent to $u(Z) \leqslant_\gamma u(x)$ for $x \in \mathbb{R}$ by (3), that is, $\frac{u(x+2)-u(x)}{2(u(x)-u(x-1))} = \gamma \frac{1+\gamma}{2} \leqslant \gamma$, where the equality follows from $\Delta_u(x) = \gamma$ for each $x \in [-2,2]$. This implies $\gamma = 1$ which yields a contradiction. $\square$

Next, we present the proof of Proposition 3, which says that (v^t)-SD has a limit of FSD, and it has a limit of SSD if only if it is one of $(1+c)_{\mathrm{A}}$-SD and $(1+r)_{\mathrm{R}}$-SD.

Proof of Proposition 3. Note that for $t > 0$,

$$\rho^{\mathrm{A}}_{v^t}(x) = -\frac{(v^t(x))''}{(v^t(x))'} = -\left(\frac{v''(x)}{v'(x)} + (t-1)\frac{v'(x)}{v(x)}\right).$$

Since $v' > 0$, it is clear that $\rho^{\mathrm{A}}_{v^t}(x) \to -\infty$ as $t \to \infty$. Using the equivalence between (1) and (2), we know that v^t-SD can be formulated via the set

$$\left\{u \in \mathscr{U} \mid u'(x) > 0,\ \rho^{\mathrm{A}}_u(x) \geqslant \rho^{\mathrm{A}}_{v^t}(x),\ x \in [a,b]\right\}.$$

Therefore, FSD is a limiting case of v^t-SD as $t \to \infty$. On the other hand, to recover SSD, it means $\rho^{\mathrm{A}}_{v^t}(x) \to 0$ as $t \downarrow t_0$, which is the oridinary differential equation

$$v''v + (t_0 - 1)(v')^2 = 0. \tag{14}$$

If $t_0 \neq 0$, then (14) is equivalent to $(v^{t_0})'' = 0$, leading to the linearity of v^{t_0}. This corresponds to the case of $(1+r)_{\mathrm{R}}$-SD, by choosing, for instance, $v(x) = x$ and $t_0 = 1$; the choice of t_0 is irrelevant here. If $t_0 = 0$, then (14) is equivalent to $(\log v)'' = 0$, leading to the linearity of $\log v$. This corresponds to the case of $(1+c)_{\mathrm{A}}$-SD, by choosing, for instance $v(x) = e^x$ and $t_0 = 0$. $\square$

A.2 Proofs of Theorems 1 and 2 in Section 3

Next, we present proofs of Theorems 1 and 2. Since Theorem 1 follows from Theorem 2 by setting $h_1 = h_2 = h$, we will show Theorem 2 below. Before proving Theorem 2, we need the following lemma, similar to Lemma 1 of Chateauneuf et al. (2005) which shows that in the definition of index of greediness, the supremum can be obtained by choosing $(x_4 - x_3)/(x_2 - x_1)$ as a fixed constant. The proof is similar and hence is omitted.

Lemma 1. *For a utility function u, $a < b$ and $\alpha > 0$, denote by $R_\alpha^{[a,b]} = \{(x_1, \dots, x_4) \in \mathbb{R}^4 : a \leqslant x_1 < x_2 \leqslant x_3 < x_4 \leqslant b,\ x_4 - x_3 = \alpha(x_2 - x_1)\}$ and $R_\alpha^{*[a,b]} = \{(x_1, \dots, x_4) \in \mathbb{R}^4 : a \leqslant x_1 < x_2 \leqslant 0 \leqslant x_3 < x_4 \leqslant b,\ x_4 - x_3 = \alpha(x_2 - x_1)\}$. Then*

$$G_u^{[a,b]} = \sup_{(x_1,\dots,x_4)\in R_\alpha^{[a,b]}} \frac{u(x_4) - u(x_3)}{x_4 - x_3} \Big/ \frac{u(x_2) - u(x_1)}{x_2 - x_1}$$

and

$$G_u^{*[a,b]} = \sup_{(x_1,\dots,x_4)\in R_\alpha^{*[a,b]}} \frac{u(x_4) - u(x_3)}{x_4 - x_3} \Big/ \frac{u(x_2) - u(x_1)}{x_2 - x_1}.$$

Proof of Theorem 2. (i) Without loss of generality we can assume $v(0) = 0$. One can verify that for any random variable X,

$$V_{u,h_1,h_2}(X) = \int_{-\infty}^{v(0)} u(v^{-1}(y))\,\mathrm{d}h_1(F_v(y)) + \int_{v(0)}^{\infty} u(v^{-1}(y))\,\mathrm{d}h_2(F_v(y)) = V_{u(v^{-1}),h_1,h_2}(v(X)),$$

where F_v is the cdf of $v(X)$. By Proposition 1, V_{u,h_1,h_2} is consistent with $\leqslant_v$ if and only if $V_{u(v^{-1}),h_1,h_2}$ is consistent with SSD. By Theorem 1 of Schmidt and Zank (2008), we have that a $\mathrm{CPT}(u, h_1, h_2)$ preference is monotone in $\leqslant_v$ if and only if $u(v^{-1}(x))$ is both concave in $x \in [0, v(0)]$ and $x \in [v(0), \infty)$, h_1, h_2 are concave, and they satisfy (10).

(ii) Assume that $\gamma > 0$ since the case of $\gamma = 0$ is trivial. We first show the necessity. Suppose for the purpose of contradiction that $\gamma G_u^{[a,0]} > P_{h_1}$, $\gamma G_u^{[0,b]} > P_{h_2}$, or $\gamma G_u^{*[a,b]} > P_{h_1,h_2}$. We shall show that a $\mathrm{CPT}(u, h_1, h_2)$ preference is not monotone in $\leqslant_\gamma^{\mathrm{S}}$ in the two cases that $\gamma G_u^{[a,0]} > P_{h_1}$ and $\gamma G_u^{*[a,b]} > P_{h_1,h_2}$, since the case $\gamma G_u^{[0,b]} > P_{h_2}$ can be shown similarly. We write $V = V_{u,h_1,h_2}$ for simplicity.

(a) If $\gamma G_u^{[a,0]} > P_{h_1}$, then from the definition of P_{h_1}, there exist $0 \leqslant p_1 < p_2 \leqslant p_3 < p_4 \leqslant 1$ such that $\gamma G_u^{[a,0]} > \frac{h_1(p_2)-h_1(p_1)}{p_2-p_1} \Big/ \frac{h_1(p_4)-h_1(p_3)}{p_4-p_3}$. Next, by Lemma 1 with $\alpha = \gamma(p_2 - p_1)/(p_4 - p_3)$, there exist $a \leqslant x_1 < x_2 \leqslant x_3 < x_4 \leqslant 0$ such that $(x_4 - x_3)(p_4 - p_3) = \gamma(x_2 - x_1)(p_2 - p_1)$ and

$$\frac{u(x_4) - u(x_3)}{u(x_2) - u(x_1)} > \frac{h_1(p_2) - h_1(p_1)}{h_1(p_4) - h_1(p_3)}. \tag{15}$$

Define two random variables X and Y such that

$$\mathbb{P}(X = x_1) = p_2,\ \mathbb{P}(X = x_2) = q - p_2,\ \mathbb{P}(X = x_3) = p_3 - q,\ \mathbb{P}(X = x_4) = 1 - p_3;$$

$$\mathbb{P}(Y = x_1) = p_1,\ \mathbb{P}(Y = x_2) = q - p_1,\ \mathbb{P}(Y = x_3) = p_4 - q,\ \mathbb{P}(Y = x_4) = 1 - p_4,$$

where $q \in [p_2, p_3]$ is a constant. We can verify that Y is obtained from X via a γ-transfer, and

$$V(X) - V(Y) = (u(x_4) - u(x_3))(h_1(p_4) - h_1(p_3)) - (u(x_2) - u(x_1))(h_1(p_2) - h_1(p_1)) > 0$$

in view of (15). This yields a contradiction. The argument in case $\gamma G_u^{[0,b]} > P_{h_2}$ is similar.

(b) If $\underline{\gamma G_u^{*[a,b]} > P_{h_1,h_2}}$, then from the definition of P_{h_1,h_2}, there exist $0 \leqslant p_1 < p_2 \leqslant p_3 < p_4 \leqslant 1$ such that $\gamma G_u^{*[a,b]} > \frac{h_1(p_2)-h_1(p_1)}{p_2-p_1} \Big/ \frac{h_2(p_4)-h_2(p_3)}{p_4-p_3}$. Next, by Lemma 1 with $\alpha = \gamma(p_2-p_1)/(p_4-p_3)$, there exist $x_1 < x_2 \leqslant 0 \leqslant x_3 < x_4$ such that $(x_4 - x_3)(p_4 - p_3) = \gamma(x_2 - x_1)(p_2 - p_1)$ and

$$\frac{u(x_4) - u(x_3)}{u(x_2) - u(x_1)} > \frac{h_1(p_2) - h_1(p_1)}{h_2(p_4) - h_2(p_3)}. \tag{16}$$

Define two random variables X and Y as in Case (a). Then we have that Y is obtained from X via a γ-transfer, and

$$V(X) - V(Y) = (u(x_4) - u(x_3))(h_2(p_4) - h_2(p_3)) - (u(x_2) - u(x_1))(h_1(p_2) - h_1(p_1)) > 0$$

in view of (16). This yields a contradiction.

Combining the above cases, we obtain the necessity statement.

Next, we show sufficiency. Without loss of generality, we assume that $u(0) = 0$. By Theorems 2.7 and 2.8 of Müller et al. (2017), it suffices to show that V is monotone in γ-transfers. Let X and Y satisfy (1) $\mathbb{P}(Y = x_t) = p_t - p_{t-1}$ for $t = 1, \ldots, n$, with $x_1 < \cdots < x_n$, $0 = p_0 \leqslant p_1 \leqslant \ldots \leqslant p_n = 1$; (2) $\mathbb{P}(X = x_t) = \mathbb{P}(Y = x_t)$ for all $t \notin \{i, k, \ell, j\}$, where $1 \leqslant i < k < \ell < j \leqslant n$; and (3) for $\eta_1 > 0$ and $\eta_2 > 0$, $\mathbb{P}(X = x_i) = p_i - p_{i-1} + \eta_1$, $\mathbb{P}(X = x_k) = p_k - p_{k-1} - \eta_1$, $\mathbb{P}(X = x_\ell) = p_\ell - p_{\ell-1} - \eta_2$, $\mathbb{P}(X = x_j) = p_j - p_{j-1} + \eta_2$, with $\gamma\eta_1(x_k - x_i) = \eta_2(x_j - x_\ell)$. Denote $s = \max\{t : x_t < 0 \leqslant x_{t+1}, t = 1, \ldots, n\}$ with $s = 0$ if $x_t < 0$ for all $t = 1, \ldots, n$, and $s = n + 1$ if

$x_t \geqslant 0$ for all $t = 1, \ldots, n$. Then it holds that

$$V(Y) = \sum_{t=1}^{s} u(x_t)(h_1(p_t) - h_1(p_{t-1})) + \sum_{t=s+1}^{n} u(x_t)(h_2(p_t) - h_2(p_{t-1})).$$

To calculate $V(X)$, we need to consider the following four subcases.

(a) $\underline{s \leqslant i}$, that is, $x_t \geqslant 0$ for any $t \in \{i, k, \ell, j\}$. In this case, we have that

$$\begin{aligned}V(X) = &\sum_{\nu=1}^{s} u(x_\nu)(h_1(p_\nu) - h_1(p_{\nu-1})) + \left(\sum_{\nu=s+1}^{i-1} + \sum_{\nu=k+1}^{\ell-1} + \sum_{\nu=j+1}^{n}\right) u(x_\nu)(h_2(p_\nu) - h_2(p_{\nu-1})) \\ &+ u(x_i)(h_2(p_i + \eta_1) - h_2(p_{i-1})) + \sum_{\nu=i+1}^{k-1} u(x_\nu)(h_2(p_\nu + \eta_1) - h_2(p_{\nu-1} + \eta_1)) \\ &+ u(x_k)(h_2(p_k) - h_2(p_{k-1} + \eta_1)) + u(x_\ell)(h_2(p_\ell - \eta_2) - h_2(p_{\ell-1})) \\ &+ u(x_j)(h_2(p_j) - h_2(p_{j-1} - \eta_2)) + \sum_{\nu=\ell+1}^{j-1} u(x_\nu)(h_2(p_\nu - \eta_2) - h_2(p_{\nu-1} - \eta_2)).\end{aligned}$$

This implies that

$$\begin{aligned}V(Y) - V(X) = &\sum_{\nu=i+1}^{k} (u(x_\nu) - u(x_{\nu-1}))(h_2(p_{\nu-1} + \eta_1) - h_2(p_{\nu-1})) \\ &- \sum_{\nu=\ell+1}^{j} (u(x_\nu) - u(x_{\nu-1}))(h_2(p_{\nu-1}) - h_2(p_{\nu-1} - \eta_2)).\end{aligned}$$

Choose k_0 and ℓ_0 such that $h_2(p_{k_0} + \eta_1) - h_2(p_{k_0}) = \min_{i \leqslant t < k}\{h_2(p_t + \eta_1) - h_2(p_t)\}$ and $h_2(p_{\ell_0}) - h_2(p_{\ell_0} - \eta_2) = \max_{\ell < t \leqslant j}\{h_2(p_{t-1}) - h_2(p_{t-1} - \eta_2)\}$. Then

$$\begin{aligned}&V(Y) - V(X) \\ &\geqslant (u(x_k) - u(x_i))(h_2(p_{k_0} + \eta_1) - h_2(p_{k_0})) - (u(x_j) - u(x_\ell))(h_2(p_{\ell_0}) - h_2(p_{\ell_0} - \eta_2)) \\ &\overset{\text{sgn}}{=} \frac{u(x_k) - u(x_i)}{x_k - x_i} \cdot \frac{h_2(p_{k_0} + \eta_1) - h_2(p_{k_0})}{\eta_1} - \gamma \frac{u(x_j) - u(x_\ell)}{x_j - x_\ell} \cdot \frac{h_2(p_{\ell_0}) - h_2(p_{\ell_0} - \eta_2)}{\eta_2} \\ &\overset{\text{sgn}}{=} \frac{h_2(p_{k_0} + \eta_1) - h_2(p_{k_0})}{\eta_1} \Big/ \frac{h_2(p_{\ell_0}) - h_2(p_{\ell_0} - \eta_2)}{\eta_2} - \gamma \frac{u(x_j) - u(x_\ell)}{x_j - x_\ell} \Big/ \frac{u(x_k) - u(x_i)}{x_k - x_i} \\ &\geqslant P_{h_2} - \gamma G_u^{[0,b]} \geqslant 0,\end{aligned}$$

where the first equality in sign $\overset{\text{sgn}}{=}$ follows from that $\gamma\eta_1(x_k - x_i) = \eta_2(x_j - x_\ell)$ and in the second $\overset{\text{sgn}}{=}$, we used the fact $u(x_k) > u(x_i)$ (otherwise, we have $u(x_k) = u(x_i)$ which implies $u(x_j) = u(x_\ell)$ or $G_u^{[0,b]} = \infty$. If $u(x_j) = u(x_\ell)$, then $V(Y) \geqslant V(X)$ holds trivially. If $G_u^{[0,b]} = \infty$, then the condition

$G_u^{[0,b]} \leqslant P_{h_2}$ implies $P_{h_2} = 1$, that is, $h_2(p) = 1_{\{p>0\}}$. It follows that $h_2(p_{\ell_0}) - h_2(p_{\ell_0} - \eta_2) = 0$ and thus, $V(Y) - V(X) \geqslant 0$.).

(b) $\underline{i < s \leqslant k}$, that is, $x_i < 0 < x_k \leqslant x_\ell < x_j$. In this case,

$$\begin{aligned}
V(X) = &\sum_{\nu=1}^{i-1} u(x_\nu)(h_1(p_\nu) - h_1(p_{\nu-1})) + \left(\sum_{\nu=k+1}^{\ell-1} + \sum_{\nu=j+1}^{n}\right) u(x_\nu)(h_2(p_\nu) - h_2(p_{\nu-1})) \\
&+ u(x_i)(h(p_i + \eta_1) - h(p_{i-1})) + \sum_{\nu=i+1}^{s} u(x_\nu)(h_1(p_\nu + \eta_1) - h_1(p_{\nu-1} + \eta_1)) \\
&+ \sum_{\nu=s+1}^{k-1} u(x_\nu)(h_2(p_\nu + \eta_1) - h_2(p_{\nu-1} + \eta_1)) \\
&+ u(x_k)(h_2(p_k) - h_2(p_{k-1} + \eta_1)) + u(x_\ell)(h_2(p_\ell - \eta_2) - h_2(p_{\ell-1})) \\
&+ u(x_j)(h_2(p_j) - h_2(p_{j-1} - \eta_2)) + \sum_{\nu=\ell+1}^{j-1} u(x_\nu)(h_2(p_\nu - \eta_2) - h_2(p_{\nu-1} - \eta_2)).
\end{aligned}$$

As a consequence,

$$\begin{aligned}
V(Y) - V(X) = &\sum_{\nu=i+1}^{s} (u(x_\nu) - u(x_{\nu-1}))(h_1(p_{\nu-1} + \eta_1) - h_1(p_{\nu-1})) \\
&- u(x_s)(h_1(p_s + \eta_1) - h_1(p_s)) + u(x_{s+1})(h_2(p_s + \eta_1) - h_2(p_s)) \\
&+ \sum_{\nu=s+2}^{k} (u(x_\nu) - u(x_{\nu-1}))(h_2(p_{\nu-1} + \eta_1) - h_2(p_{\nu-1})) \\
&- \sum_{\nu=\ell+1}^{j} (u(x_\nu) - u(x_{\nu-1}))(h_2(p_{\nu-1}) - h_2(p_{\nu-1} - \eta_2)).
\end{aligned}$$

Choose k_1, k_2 and ℓ_0 such that $h_1(p_{k_1} + \eta_1) - h_1(p_{k_1}) = \min_{i \leqslant t \leqslant s}\{h_1(p_t + \eta_1) - h_1(p_t)\}$, $h_2(p_{k_2} + \eta_1) - h_2(p_{k_2}) = \min_{s \leqslant t < k}\{h_2(p_t + \eta_1) - h_2(p_t)\}$ and $h_2(p_{\ell_0}) - h_2(p_{\ell_0} - \eta_2) = \max_{\ell < t \leqslant j}\{h_2(p_{t-1}) - h_2(p_{t-1} - \eta_2)\}$. Then we have

$$\begin{aligned}
&V(Y) - V(X) \\
&\geqslant -u(x_i)(h_1(p_{k_1} + \eta_1) - h_1(p_{k_1})) + u(x_k)(h_2(p_{k_2} + \eta_1) - h_2(p_{k_2})) \\
&\quad - (u(x_j) - u(x_\ell))(h_2(p_{\ell_0}) - h_2(p_{\ell_0} - \eta_2)) \\
&\overset{\text{sgn}}{=} \frac{-u(x_i)}{x_k - x_i} \cdot \frac{h_1(p_{k_1} + \eta_1) - h_1(p_{k_1})}{\eta_1} + \frac{u(x_k)}{x_k - x_i} \cdot \frac{h_2(p_{k_2} + \eta_1) - h_2(p_{k_2})}{\eta_1} \\
&\quad - \gamma \, \frac{u(x_j) - u(x_\ell)}{x_j - x_\ell} \cdot \frac{h_2(p_{\ell_0}) - h_2(p_{\ell_0} - \eta_2)}{\eta_2}
\end{aligned}$$

$$= \lambda \left(\frac{u(0) - u(x_i)}{0 - x_i} \cdot \frac{h_1(p_{k_1} + \eta_1) - h_1(p_{k_1})}{\eta_1} - \gamma \frac{u(x_j) - u(x_\ell)}{x_j - x_\ell} \cdot \frac{h_2(p_{\ell_0}) - h_2(p_{\ell_0} - \eta_2)}{\eta_2} \right)$$
$$+ (1-\lambda) \left(\frac{u(x_k) - u(0)}{x_k - 0} \cdot \frac{h_2(p_{k_2} + \eta_1) - h_2(p_{k_2})}{\eta_1} - \gamma \frac{u(x_j) - u(x_\ell)}{x_j - x_\ell} \cdot \frac{h_2(p_{\ell_0}) - h_2(p_{\ell_0} - \eta_2)}{\eta_2} \right)$$
$$=: \lambda I_1 + (1 - \lambda) I_2, \tag{17}$$

where $\lambda = -x_i/(x_k - x_i) > 0$. Here, the first inequality follows from $u(x_{s+1}) \geqslant u(0) = 0$ and $u(x_s) \leqslant u(0) = 0$. Note that

$$I_2 = \frac{u(x_k) - u(0)}{x_k - 0} \cdot \frac{h_2(p_{k_2} + \eta_1) - h_2(p_{k_2})}{\eta_1} - \gamma \frac{u(x_j) - u(x_\ell)}{x_j - x_\ell} \cdot \frac{h_2(p_{\ell_0}) - h_2(p_{\ell_0} - \eta_2)}{\eta_2}$$
$$\overset{\text{sgn}}{=} \frac{(h_2(p_{k_2} + \eta_1) - h_2(p_{k_2}))}{\eta_1} \Big/ \frac{h_2(p_{\ell_0}) - h_2(p_{\ell_0} - \eta_2)}{\eta_2} - \gamma \frac{u(x_j) - u(x_\ell)}{x_j - x_\ell} \Big/ \frac{u(x_k) - u(0)}{x_k - 0}$$
$$\geqslant P_{h_2} - \gamma G_u^{[0,b]} \geqslant 0,$$

where the first inequality follows from that $0 \leqslant x_k < x_\ell < x_j$. We also used $u(x_k) > u(0)$ for $\overset{\text{sgn}}{=}$. This is because if $u(x_k) = u(0)$, then one can easily verify $V(X) \leqslant V(Y)$ by similar arguments in the bracket of case (a). In addition,

$$I_1 = \frac{u(0) - u(x_i)}{0 - x_i} \cdot \frac{h_1(p_{k_1} + \eta_1) - h_1(p_{k_1})}{\eta_1} - \gamma \frac{u(x_j) - u(x_\ell)}{x_j - x_\ell} \cdot \frac{h_2(p_{\ell_0}) - h_2(p_{\ell_0} - \eta_2)}{\eta_2}$$
$$\overset{\text{sgn}}{=} \frac{(h_1(p_{k_1} + \eta_1) - h_1(p_{k_1}))}{\eta_1} \Big/ \frac{h_2(p_{\ell_0}) - h_2(p_{\ell_0} - \eta_2)}{\eta_2} - \gamma \frac{u(x_j) - u(x_\ell)}{x_j - x_\ell} \Big/ \frac{u(0) - u(x_i)}{0 - x_i}$$
$$\geqslant P_{h_1,h_2} - \gamma G_u^{*[a,b]} \geqslant 0,$$

where the first inequality follows from that $x_i < 0 \leqslant x_\ell < x_j$. Substituting I_1 and I_2 into (17) yields that $V(Y) - V(X) \geqslant 0$.

(c) $\underline{k \leqslant s < \ell}$, that is, $x_i < x_k \leqslant 0 \leqslant x_\ell < x_j$. In this case,

$$V(X) = \left(\sum_{\nu=1}^{i-1} + \sum_{\nu=k+1}^{s} \right) u(x_\nu)(h_1(p_\nu) - h_1(p_{\nu-1})) + \left(\sum_{\nu=s+1}^{\ell-1} + \sum_{\nu=j+1}^{n} \right) u(x_\nu)(h_2(p_\nu) - h_2(p_{\nu-1}))$$
$$+ u(x_i)(h_1(p_i + \eta_1) - h_1(p_{i-1})) + \sum_{\nu=i+1}^{k-1} u(x_\nu)(h_1(p_\nu + \eta_1) - h_1(p_{\nu-1} + \eta_1))$$
$$+ u(x_k)(h_1(p_k) - h_1(p_{k-1} + \eta_1)) + u(x_\ell)(h_2(p_\ell - \eta_2) - h_2(p_{\ell-1}))$$
$$+ u(x_j)(h_2(p_j) - h_2(p_{j-1} - \eta_2)) + \sum_{\nu=\ell+1}^{j-1} u(x_\nu)(h_2(p_\nu - \eta_2) - h_2(p_{\nu-1} - \eta_2)).$$

As a consequence,

$$V(Y) - V(X) = \sum_{\nu=i+1}^{k} (u(x_\nu) - u(x_{\nu-1}))(h_1(p_{\nu-1} + \eta_1) - h_1(p_{\nu-1})) - \sum_{\nu=\ell+1}^{j} (u(x_\nu) - u(x_{\nu-1}))(h_2(p_{\nu-1}) - h_2(p_{\nu-1} - \eta_2)).$$

Choose k_0 and ℓ_0 such that $h_1(p_{k_0} + \eta_1) - h_1(p_{k_0}) = \min_{i \leqslant t < k}\{h_1(p_t + \eta_1) - h_1(p_t)\}$ and $h_2(p_{\ell_0}) - h_2(p_{\ell_0} - \eta_2) = \max_{\ell \leqslant t < j}\{h_2(p_t) - h_2(p_t - \eta_2)\}$. Then we have

$$\begin{aligned}
&V(Y) - V(X) \\
&\geqslant (u(x_k) - u(x_i))(h_1(p_{k_0} + \eta_1) - h_1(p_{k_0})) - (u(x_j) - u(x_\ell))(h_2(p_{\ell_0}) - h_2(p_{\ell_0} - \eta_2)) \\
&\overset{\text{sgn}}{=} \frac{u(x_k) - u(x_i)}{x_k - x_i} \cdot \frac{h_1(p_{k_0} + \eta_1) - h_1(p_{k_0})}{\eta_1} - \gamma \frac{u(x_j) - u(x_\ell)}{x_j - x_\ell} \cdot \frac{h_2(p_{\ell_0}) - h_2(p_{\ell_0} - \eta_2)}{\eta_2} \\
&\overset{\text{sgn}}{=} \frac{h_1(p_{k_0} + \eta_1) - h_1(p_{k_0})}{\eta_1} \Big/ \frac{h_2(p_{\ell_0}) - h_2(p_{\ell_0} - \eta_2)}{\eta_2} - \gamma \frac{u(x_j) - u(x_\ell)}{x_j - x_\ell} \Big/ \frac{u(x_k) - u(x_i)}{x_k - x_i} \\
&\geqslant P_{h_1,h_2} - \gamma G_u^{*[a,b]} \geqslant 0.
\end{aligned}$$

Here, we used the fact $u(x_k) > u(x_i)$ for the second $\overset{\text{sgn}}{=}$, since otherwise $V(Y) - V(X) \geqslant 0$ holds trivially.

(d) $\underline{\ell \leqslant s < j \text{ or } j \leqslant s, \text{ that is, } x_i < x_k \leqslant x_\ell < 0 < x_j \text{ or } x_i < x_k \leqslant x_\ell < x_j \leqslant 0.}$ These two cases can be proved by similar arguments to those in cases (a) and (b). Hence, the details are omitted here.

Combining cases (a)-(d), we conclude that $V(X) \leqslant V(Y)$. Thus, we complete the proof. □

A.3 Proofs of results in Section 4

We proceed to prove results in Section 4 on properties of the few notions of fractional SD. We first show Proposition 4.

Proof of Proposition 4. The $\Leftarrow$ implication is trivial because $s(0) = r$, and we consider the $\Rightarrow$ implication below. Denote $X_0 = X^{1/r}$ and $Y_0 = Y^{1/r}$. Then $X \leqslant_r^{\mathrm{R}} Y$ if and only if $X_0 \leqslant_{\mathrm{SSD}} Y_0$. We aim to find the smallest value of s such that $X + \beta \leqslant_s^{\mathrm{R}} Y + \beta$, that is, $(X+\beta)^{1/s} \leqslant_{\mathrm{SSD}} (Y+\beta)^{1/s}$. Hence, we need

$$v(x) := (x^r + \beta)^{1/s}, \quad x \in [a^{1/r}, b^{1/r}]$$

to be concave which is equivalent to

$$v''(x) = \frac{r}{s}(r-1)(x^r+\beta)^{1/s-1}x^{r-2} + \frac{r^2}{s}\left(\frac{1}{s}-1\right)(x^r+\beta)^{1/s-2}x^{2r-2}$$
$$\overset{\text{sgn}}{=} (r-1)(x^r+\beta) + r\left(\frac{1}{s}-1\right)x^r \leqslant 0,$$

that is,

$$\frac{1}{s} \leqslant 1 + \frac{1-r}{r} \times \frac{x^r+\beta}{x^r}.$$

Note that if $\beta \geqslant 0$, then

$$\frac{x^r+\beta}{x^r} \geqslant 1 + \frac{\beta}{b};$$

Then v is concave on $[a^{1/r}, b^{1/r}]$ if $s \geqslant \frac{rb}{b+\beta-r\beta}$. If $\beta \leqslant 0$, then

$$\frac{x^r+\beta}{x^r} \geqslant 1 + \frac{\beta}{a}$$

and v is concave on $[a^{1/r}, b^{1/r}]$ if $s \geqslant \frac{ra}{a+\beta-r\beta}$. □

Below, we discuss how $X \leqslant_{\mathrm{SSD}} Y$ implies $(1+c)_{\mathrm{A}}$-SD and $(1+\gamma)_{\mathrm{S}}$-SD for a small modification of Y. It turns out that for bounded X and Y, both $X \leqslant^{\mathrm{A}}_{c} Y+\varepsilon$ and $X \leqslant^{\mathrm{S}}_{\gamma} Y+\varepsilon$ hold for some c and γ close to 1, which will be used in the proof of Theorem 3.

Lemma 2. *Suppose that X and Y are two random variables such that $X \leqslant_{\mathrm{SSD}} Y$ and $|X|, |Y| \leqslant m$ for some $m > 0$. For any $\varepsilon \in (0, m]$, $c \geqslant 1 - \frac{2\varepsilon}{2\varepsilon+em^2} \in [0,1)$, $\gamma \geqslant 1 - \frac{2\varepsilon}{\varepsilon+m} \in [0,1)$, and $r \geqslant \frac{m^2}{m^2+2\varepsilon} \in [0,1)$, we have $X \leqslant^{\mathrm{A}}_{c} Y+\varepsilon$, $X \leqslant^{\mathrm{S}}_{\gamma} Y+\varepsilon$ and $X \leqslant^{\mathrm{R}}_{r} Y+\varepsilon$.*

Proof. We first recall the definition of γ-transfer given by Müller et al. (2017). For two discrete random variables X and Y, we say that X is obtained from Y via a *γ-transfer* if there exist two random variables $\hat{X}$, $\hat{Y}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $\omega_1, \omega_2 \in \Omega$, $x_1 < x_2 \leqslant x_3 < x_4$ and $\eta_1, \eta_2 > 0$ with $\eta_2(x_4 - x_3) = \gamma\eta_1(x_2 - x_1)$, $\mathbb{P}(\{\omega_i\}) = \eta_i$, $i = 1, 2$, such that $\hat{X} \overset{\mathrm{d}}{=} X$, $\hat{Y} \overset{\mathrm{d}}{=} Y$, where $\overset{\mathrm{d}}{=}$ represents equality in distribution, and

$$\hat{X}(\omega_1) = x_1,\ \hat{X}(\omega_2) = x_4,\ \hat{Y}(\omega_1) = x_2,\ \hat{Y}(\omega_2) = x_3 \text{ and } \hat{X}(\omega) = \hat{Y}(\omega) \text{ for } \omega \neq \omega_1, \omega_2. \tag{18}$$

It is obvious that $X \leqslant^{\mathrm{S}}_{\gamma} Y$.

Let X and Y be two random variables such that $X \leqslant_{\mathrm{SSD}} Y$. To find the constants $\gamma_{\varepsilon,m}$ and $c_{\varepsilon,m}$, it suffices consider the case that X is obtained from Y via a 1-transfer. The general case can be argued as a limit of 1-transfers; see e.g., Theorems 2.7 and 2.8 of Müller et al. (2017). Without loss of generality, assume $X = \hat{X}$ and $Y = \hat{Y}$ satisfy (18) with $\gamma = 1$. Define Y_ε as

$$Y_\varepsilon(\omega_1) = x_2 + \varepsilon, \quad Y_\varepsilon(\omega_2) = x_3 + \varepsilon, \quad Y_\varepsilon(\omega) = Y(\omega), \quad \omega \neq \omega_1, \omega_2,$$

It is obvious that $Y_\varepsilon \leqslant_{\mathrm{FSD}} Y + \varepsilon$. Denote $s = x_2 - x_1$ and $t = x_4 - x_3$. Then $t\eta_2 = s\eta_1$ and $s + t \leqslant 2m$. Let us compare X and Y_ε.

(i) We first consider $(1+\gamma)_{\mathrm{S}}$-SD. Noting that

$$\frac{\eta_2(t-\varepsilon)}{\eta_1(s+\varepsilon)} = \frac{(t-\varepsilon)s}{t(s+\varepsilon)} = \frac{1-\varepsilon/t}{1+\varepsilon/s} \leqslant \frac{m-\varepsilon}{m+\varepsilon} =: \gamma_\varepsilon, \tag{19}$$

where the first equality follows from $\eta_2 t = \eta_1 s$. It is obvious that (19) implies $X \leqslant_{\gamma_\varepsilon} Y_\varepsilon$, and thus $X \leqslant_{\gamma_\varepsilon} Y + \varepsilon$.

(ii) We next consider $(1+c)_{\mathrm{A}}$-SD. By Theorem 2 (i), we need to find a $\lambda > 0$ such that $e^{\lambda X} \leqslant_{\mathrm{SSD}} e^{\lambda Y_\varepsilon}$, that is,

$$\frac{\eta_2(e^{\lambda x_4} - e^{\lambda(x_3+\varepsilon)})}{\eta_1(e^{\lambda(x_2+\varepsilon)} - e^{\lambda x_1})} = \frac{s(e^{\lambda x_4} - e^{\lambda(x_3+\varepsilon)})}{t(e^{\lambda(x_2+\varepsilon)} - e^{\lambda x_1})} \leqslant 1, \tag{20}$$

where the first inequality follows from $\eta_2 t = \eta_1 s$, or equivalently,

$$\frac{e^{\lambda x_4} - e^{\lambda(x_3+\varepsilon)}}{x_4 - x_3} \leqslant \frac{e^{\lambda(x_2+\varepsilon)} - e^{\lambda x_1}}{x_2 - x_1}.$$

For this, it suffices to make the left-hand side less or equal to λ and the right-hand side greater or equal to λ; thus

$$e^{\lambda x_4} - e^{\lambda(x_3+\varepsilon)} \leqslant \lambda(x_4 - x_3), \tag{21}$$

and

$$e^{\lambda(x_2+\varepsilon)} - e^{\lambda x_1} \geqslant \lambda(x_2 - x_1). \tag{22}$$

We first deal with (21). Noting that $e^{\lambda(x_3+\varepsilon)} \geqslant 1 + \lambda x_3 + \lambda\varepsilon$, it suffices to show

$$e^{\lambda x_4} \leqslant \lambda(x_4 - x_3) + 1 + \lambda x_3 + \lambda\varepsilon = 1 + \lambda x_4 + \lambda\varepsilon. \tag{23}$$

Note that for $\lambda \leqslant 1/m$, we have $e^{\lambda x_4} \leqslant e$. Using Langrange's formula, we have

$$e^{\lambda x_4} \leqslant 1 + \lambda x_4 + \frac{e}{2}(\lambda x_4)^2 \leqslant 1 + \lambda x_4 + \frac{e}{2}\lambda^2 m^2.$$

Hence, choosing $\lambda \leqslant \frac{2\varepsilon}{em^2}$ is sufficient for (23). Note that since $\varepsilon \leqslant m$, the condition $\lambda \leqslant 1/m$ is automatic. Similarly, (22) holds for such choice of λ. Thus, $\lambda \leqslant \frac{2\varepsilon}{em^2}$ is sufficient for (20).

This corresponds to $c \geqslant 1 - \frac{2\varepsilon}{2\varepsilon + em^2}$.

(iii) At last we consider $(1+r)_{\mathrm{R}}$-SD. Note that for $r \geqslant m^2/(m^2+2\varepsilon)$, by Taylor's expansion, we have

$$x_4^{1/r} - (x_3+\varepsilon)^{1/r} \leqslant \frac{x_4 - x_3}{r} \quad \text{and} \quad (x_2+\varepsilon)^{1/r} - x_1^{1/r} \geqslant \frac{x_2 - x_1}{r}.$$

which is equivalent to

$$\frac{\eta_2(x_4^{1/r} - (x_3+\varepsilon)^{1/r})}{\eta_1((x_2+\varepsilon)^{1/r} - x_1^{1/r})} = \frac{s(x_4^{1/r} - (x_3+\varepsilon)^{1/r})}{t((x_2+\varepsilon)^{1/r} - x_1^{1/r})} \leqslant 1, \tag{24}$$

where the first inequality follows from $\eta_2 t = \eta_1 s$. By Theorem 2 (i), we know (24) is equivalent to $X^{1/r} \leqslant_{\mathrm{SSD}} (Y+\varepsilon)^{1/r}$, that is, $X \leqslant_r^{\mathrm{R}} Y + \varepsilon$. □

Proof of Theorem 3. (i) The "if" direction is trivial, and below we show the "only-if" direction for $c \in (0,1)$. Note that if $X \leqslant_c^{\mathrm{A}} Y$, then $\alpha X \leqslant_{\alpha c/(1-c+\alpha c)}^{\mathrm{A}} \alpha Y$ for all $\alpha > 0$. Hence, if $\prec$ is monotone in $\leqslant_c^{\mathrm{A}}$ for some $c \in (0,1)$, then it is also monotone in $\leqslant_{c'}^{\mathrm{A}}$ for all $c' \in (0,1)$. For any X and Y such that $X \leqslant_{\mathrm{SSD}} Y$, we aim to show $X \preccurlyeq Y$. We consider the following two cases. By Lemma 2, for $n \geqslant \mathbb{N}$, there exists $c_n \in (0,1)$ such that $X \leqslant_{c_n}^{\mathrm{S}} Y_n := Y + 1/n$ and hence $X \preccurlyeq Y_n$, $n \in \mathbb{N}$. Then by $|Y_n - Y| \to 0$ uniformly and the upper semi-continuity of the preference $\preccurlyeq$, we have $X \preccurlyeq Y$.

Note that $X \leqslant_{c_n}^{\mathrm{S}} Y_n$ is equivalent to $X_n := X - 1/n \leqslant_{c_n}^{\mathrm{S}} Y$ and hence $X_n \preccurlyeq Y$, $n \in \mathbb{N}$. Then by $|X_n - X| \to 0$ uniformly and the lower semi-continuity of the preference $\preccurlyeq$, we have $X \preccurlyeq Y$.

(ii) Similarly, we only need to show the "only-if" direction. We first show that $\preccurlyeq$ is monotone in $\leqslant_r^{\mathrm{R}}$ if and only if $\preccurlyeq$ is monotone in $\leqslant_{r'}^{\mathrm{R}}$ for $r', r \in (0,1)$. It suffices to show the necessity for $0 < r' < r < 1$. Note that for any $X, Y \in \mathcal{X}$ such that $X \leqslant_r^{\mathrm{R}} Y$, let $d = \max\{\text{ess-sup}X, \text{ess-sup}Y\} < \infty$. By Proposition 4, there exists $\beta = (r - r')d/(r'(1-r)) > 0$ such that $X + \beta \leqslant_{r'}^{\mathrm{R}} Y + \beta$. The remaining proof is similar to (i). □

Proof of Proposition 5. (i) Note that for any $u \in \mathscr{U}_\gamma^{\mathrm{S}}$, if it is not a constant function, then there exists y such that $u'(y) = \delta > 0$. Hence, for any $x < y$, $u'(x) \geqslant \gamma u'(y) = \gamma\delta$. This implies $u(x) \leqslant u(y) + \gamma\delta(y - x)$ for $x \leqslant y$. Hence,

$$\mathbb{E}[u(X) I_{\{X \leqslant y\}}] \leqslant \mathbb{E}[(u(y) + \gamma\delta(y - X) I_{\{X \leqslant y\}}] = -\infty.$$

Thus, $\mathbb{E}[u(X)] \leqslant \mathbb{E}[u(Y)]$ provided the expectations exist, that is, (i) follows.

(ii) It suffices to consider the case $X \leqslant 0$ as the partial order SSD is closed under mixture. For $\lambda = 1/c - 1$ and $\lambda' = 1/c' - 1$, define $y = \frac{1}{\lambda}\log \mathbb{E}[e^{\lambda X}]$. We can verify that $e^{\lambda X} \leqslant_{\mathrm{SSD}} e^{\lambda y}$, that is, $X \leqslant_c^{\mathrm{A}} y$. Note that for non-degenerated random variable X, $\frac{1}{\lambda'}\log \mathbb{E}[e^{\lambda' X}] > \frac{1}{\lambda}\log \mathbb{E}[e^{\lambda X}] = y$. This implies $e^{\lambda' X} \not\leqslant_{\mathrm{SSD}} e^{\lambda' y}$, that is, $X \not\leqslant^{\mathrm{S}}_{c'} Y$ for $Y = y$. $\square$

Proof of Proposition 6. Suppose that $\mathbb{P}(X < d) > 0$. Take $u(x) = \min\{x, d\}$, $x \in \mathbb{R}$. It is straightforward to see that $d \not\leqslant_{\mathrm{SSD}} X$ and thus, $d \not\leqslant^{\mathrm{A}}_{c} X$ for any $c \in [0, 1]$. On the other hand, $\mathbb{E}[e^{\lambda X}] = \infty$ implies $e^{\lambda X} \not\leqslant_{\mathrm{SSD}} d$ and thus, X is not dominated by d in $(1+c)_{\mathrm{A}}$-SD by Theorem 2. $\square$

Details of Example 4. (1) X_k has the distribution

$$F_k(x) = \Phi\left(\frac{\log x - \mu_k}{\sigma_k}\right), \quad x > 0,$$

where Φ is the standard normal distribution. We can check that F_1 and F_2 are single crossing at point e^{t_0} with $t_0 = \frac{\mu_1\sigma_2 - \mu_2\sigma_1}{\sigma_2 - \sigma_1}$. The the value of $\gamma_{\min}$ follows from the characterization of $(1+\gamma)_{\mathrm{S}}$-SD of Theorem 2.4 of Müller et al. (2017). By Theorem 2, if $c_{\min} \in (0, 1)$, then $c_{\min} = 1/(1 + \lambda_{\max})$ with $\lambda_{\max}$ the largest value of λ such that $e^{\lambda X_1} \leqslant_{\mathrm{SSD}} e^{\lambda X_2}$. Then $\lambda_{\max}$ is the unique solution to the following equation

$$\int_0^{e^{t_0}} e^{\lambda x} \left(F_1(x) - F_2(x)\right) \mathrm{d}x = \int_{e^{t_0}}^{\infty} e^{\lambda x} \left(F_2(x) - F_1(x)\right) \mathrm{d}x. \tag{25}$$

Denote by

$$h_\lambda(x) := e^{\lambda e^x} \left(\Phi\left(\frac{x - \mu_2}{\sigma_2}\right) - \Phi\left(\frac{x - \mu_1}{\sigma_1}\right)\right).$$

We have

$$\int_{e^{t_0}}^{\infty} e^{\lambda x} \left(F_2(x) - F_1(x)\right) \mathrm{d}x = \int_{t_0}^{\infty} h_\lambda(x) \mathrm{d}x.$$

Note that for any $\lambda > 0$ by L'Hôpital's principle, we have

$$\lim_{x \to \infty} \frac{h_\lambda(x)}{x} = \infty,$$

which implies that the integration of the right hand side of (25) equals to infinity. The left hand side is always finite, and thus, λ satisfying (25) does not exist. Hence $X_1 \not\leqslant^{\mathrm{A}}_{c} X_2$ for any $c < 1$.

(2) First note that $(1 + x/\mu)^{-\alpha}$ is decreasing in $\alpha > 0$ and increasing in $\mu > 0$ and thus, when $\alpha_1 \geqslant \alpha_2$ and $\mu_1 \leqslant \mu_2$,

$$1 - F_1(x) = \left(1 + \frac{x}{\mu_1}\right)^{-\alpha_1} \leqslant \left(1 + \frac{x}{\mu_2}\right)^{-\alpha_2} = 1 - F_2(x),$$

that is, $F_1 \leqslant_{\mathrm{FSD}} F_2$. Hence, without loss of generality we assume $\alpha_1 > \alpha_2$ and $\mu_1 \geqslant \mu_2$. Denote

$$g(x) = \frac{(x + \mu_1)^{\alpha_1}}{(x + \mu_2)^{\alpha_2}}, \quad x \geqslant 0.$$

Note that $g(x) = \mu_1^{\alpha_1}/\mu_2^{\alpha_2}$ if and only if $F_1(x) = F_2(x)$; and

$$\frac{\mathrm{d}}{\mathrm{d}x} \log g(x) = \frac{\alpha_1}{x+\mu_1} - \frac{\alpha_2}{x+\mu_2} = \frac{(\alpha_1 - \alpha_2)x + \alpha_1\mu_2 - \alpha_2\mu_1}{(x+\mu_1)(x+\mu_2)}, \quad x > 0.$$

This implies $\frac{\mathrm{d}^2}{\mathrm{d}x^2} \log g(x) > 0$ and thus, $g(x)$ is convex in $x > 0$. We then consider two cases.

(i) If $\alpha_1\mu_2 - \alpha_2\mu_1 \geqslant 0$, then $\mathrm{d} \log g(x)/\mathrm{d}x > 0$ and thus, $g(x)$ is increasing. Then by $g(0) = \mu_1^{\alpha_1}/\mu_2^{\alpha_2}$, we have $g(x) > \mu_1^{\alpha_1}/\mu_2^{\alpha_2}$, that is, $1 - F_1(x) \leqslant 1 - F_2(x)$ for all $x \geqslant 0$. Then we have $F_1 \leqslant_{\mathrm{FSD}} F_2$.

(ii) If $\alpha_1\mu_2 - \alpha_2\mu_1 < 0$, then $\mathrm{d} \log g(x)/\mathrm{d}x$ is negative for $x < (\alpha_2\mu_1 - \alpha_1\mu_2)/(\alpha_2 - \alpha_1)$ and then positive when $x > (\alpha_2\mu_1 - \alpha_1\mu_2)/(\alpha_2 - \alpha_1)$. Note that $g(0) = \mu_1^{\alpha_1}/\mu_2^{\alpha_2}$ and $\lim_{x\to\infty} g(x) = \infty$. There exists unique $x_0 > 0$ such that $g(x_0) = \mu_1^{\alpha_1}/\mu_2^{\alpha_2}$, that is, $F_1(x_0) = F_2(x_0)$. That is, F_1 and F_2 are single crossing at x_0. For $x < x_0$, we have $g(x) < \mu_1^{\alpha_1}/\mu_2^{\alpha_2}$, that is,

$$\frac{\mu_1^{\alpha_1}}{(x+\mu_1)^{\alpha_1}} > \frac{\mu_2^{\alpha_2}}{(x+\mu_2)^{\alpha_2}}, \quad x < x_0.$$

This is equivalent to $F_1(x) < F_2(x)$ for $x < x_0$ and $F_1(x) > F_2(x)$ for $x > x_0$ (this implies $F_1 \not\leqslant_{\mathrm{SSD}} F_2$). Thus, if $1 \geqslant \alpha_1 > \alpha_2$ or $\mathbb{E}^{F_1}[X_1] = \mu_1/(\alpha_1 - 1) < \mathbb{E}^{F_2}[X_2] = \mu_2/(\alpha_2 - 1)$ ($\alpha_1 > \alpha_2 > 1$), we have $F_2 \not\leqslant_{\mathrm{SSD}} F_1$.

We only need to consider the case $\alpha_1 > \alpha_2 > 1$ and $\mathbb{E}^{F_1}[X_1] \geqslant \mathbb{E}^{F_2}[X_2]$ in which case we have $F_2 \leqslant_{\mathrm{SSD}} F_1$, then by the characterization of $(1+\gamma)_{\mathrm{S}}$-SD of Theorem 2.4 of Müller et al. (2017), we have $F_2 \leqslant_\gamma^{\mathrm{S}} F_1$ with $\gamma \geqslant \gamma_{\min}$

$$\gamma_{\min} = \frac{\int_{x_0}^\infty F_1(x) - F_2(x)\,\mathrm{d}x}{\int_0^{x_0} F_2(x) - F_1(x)\,\mathrm{d}x} = \frac{\int_{x_0}^\infty \frac{\mu_2^{\alpha_2}}{(x+\mu_2)^{\alpha_2}} - \frac{\mu_1^{\alpha_1}}{(x+\mu_1)^{\alpha_1}}\,\mathrm{d}x}{\int_0^{x_0} \frac{\mu_1^{\alpha_1}}{(x+\mu_1)^{\alpha_1}} - \frac{\mu_2^{\alpha_2}}{(x+\mu_2)^{\alpha_2}}\,\mathrm{d}x} \in (0,1]$$

In contrast, for any $\alpha_1 > \alpha_2 > 0$, we have $\int_{x_0}^\infty (F_1(\log x/\lambda) - F_2(\log x/\lambda))\,\mathrm{d}x = \lambda \int_{x_0}^\infty e^{\lambda x}(F_1(x) - F_2(x))\,\mathrm{d}x = \infty$ which implies $F_2 \not\leqslant_c^{\mathrm{A}} F_1$ for $c \in (0,1)$. □

A.4 Proofs of results in Section 5

This section contains proofs of Propositions 7, 8 and 9.

Proof of Proposition 7. By Theorem 2 (i), we have $\lambda_{\max}$ is the maximum value of λ such that

$$e^{\lambda X} \leqslant_{\mathrm{SSD}} e^{\lambda Y}. \tag{26}$$

By that (26) is equivalent to $\mathbb{E}[e^{\lambda X_p}] \leqslant \mathbb{E}[e^{\lambda Y_p}]$ for $p \in (0,1)$, then (13) follows immediately. We then consider the case that F single crosses G. In this case, we have the distribution of $e^{\lambda X}$ single crosses that of $e^{\lambda Y}$ and thus, we have (26) holds if and only if $\mathbb{E}[e^{\lambda X}] \leqslant \mathbb{E}[e^{\lambda Y}]$. Note that $\lambda \mapsto \mathbb{E}[e^{\lambda X}]$ is continuous in λ and thus, we have $\lambda_{\max}$ is the largest value satisfying $\mathbb{E}[e^{\lambda X}] = \mathbb{E}[e^{\lambda Y}]$. We next assert that for any $\lambda < \lambda_{\max}$, $\mathbb{E}[e^{\lambda X}] < \mathbb{E}[e^{\lambda Y}]$. By Theorem 3.A.4 of Shaked and Shanthikumar (2007), there exist $\hat{X}$ and $\hat{Y}$ such that $\hat{X} \overset{\text{st}}{=} e^{\lambda_{\max} X}$, $\hat{Y} \overset{\text{st}}{=} e^{\lambda_{\max} Y}$ and

$$\mathbb{E}[\hat{X}|\hat{Y}] = \hat{Y} \quad \text{a.s.}$$

Then $e^{\lambda X} \overset{\text{st}}{=} \hat{X}^\delta$ and $e^{\lambda Y} \overset{\text{st}}{=} \hat{Y}^\delta$ with $\delta := \lambda/\lambda_{\max} < 1$, and

$$\mathbb{E}[\hat{X}^\delta] = \mathbb{E}[\mathbb{E}[\hat{X}^\delta|\hat{Y}^\delta]] = \mathbb{E}[\mathbb{E}[\hat{X}^\delta|\hat{Y}]] \leqslant \mathbb{E}[(\mathbb{E}[\hat{X}|\hat{Y}])^\delta] = \mathbb{E}[\hat{Y}^\delta],$$

where the inequality follows from $x \mapsto x^\delta$ is a strictly concave function. Note that $\mathbb{E}[\hat{X}^\delta|\hat{Y}] \leqslant \mathbb{E}[\hat{X}|\hat{Y}])^\delta]$ a.s., and $\mathbb{P}(\mathbb{E}[\hat{X}^\delta|\hat{Y}] = (\mathbb{E}[\hat{X}|\hat{Y}])^\delta) < 1$. Thus, we have $\mathbb{E}[\hat{X}^\delta] < \mathbb{E}[\hat{Y}^\delta]$, that is, $\mathbb{E}[e^{\lambda X}] < \mathbb{E}[e^{\lambda Y}]$. It then follows that $\lambda_{\max}$ is the unique solution satisfying the equation. □

Proof of Proposition 8. It suffices to consider $n = m$ by replacing m and n with $m \wedge n$. Since F and G are continuous and strictly single-crossing, we know that the probability that F_n and G_n are single-crossing tends to 1. Hence, we only need to consider $\hat{c}_{\min}$ when F_n and G_n are single-crossing.

Define the random function $g_n(\lambda) = \int e^{\lambda x}\,\mathrm{d}F_n(x) - \int e^{\lambda x}\,\mathrm{d}G_n(x)$ for $\lambda > 0$, and let $g(\lambda) = \int e^{\lambda x}\,\mathrm{d}F(x) - \int e^{\lambda x}\,\mathrm{d}G(x)$ for $\lambda > 0$. Since $x \mapsto e^{\lambda x}$ is a bounded continuous function, we know that $g_n(\lambda) \to g(\lambda)$ by the continuous mapping theorem. We also note that $g(\lambda) = 0$ has a unique root $\lambda_{\max} > 0$, and $g(\lambda)$ is bounded away from 0 outside a neighbourhood of $\lambda_{\max}$ since for all λ, $g(\lambda) > 0$ implies $g'(\lambda) > 0$ and $g(\lambda) < 0$ implies $g'(\lambda) < 0$. Therefore, the unique root λ_n of $g_n(\lambda) = 0$, which exists with probability tending to 1, converges to λ with probability 1. This shows $\hat{c}_{\min} \to c_{\min}$ in probability. □

Proof of Proposition 9. As $\gamma \leqslant 1$, it suffices to show that $\tilde{\gamma}_{\min} \to \gamma_{\min}$ in probability as $n, m \to \infty$. Note that $|x_+ - y_+| \leqslant |x - y|$ for each $x, y \in \mathbb{R}$. It follows that

$$\left|\int_{\ell_F}^t (G_m(x) - F_n(x))_+ - (G(x) - F(x))_+ \mathrm{d}x\right| \leqslant \int_{\ell_F}^t |(G_m(x) - G(x)) - (F_n(x) - F(x))|\,\mathrm{d}x.$$

Note that by the assumption on empirical distribution and the underlying distribution is continuous, we have $\sup_{x\in\mathbb{R}} |G_m(x) - G(x)| \to 0$ and $\sup_{x\in\mathbb{R}} |F_n(x) - F(x))| \to 0$ as $m, n \to \infty$ and thus,

$$\sup_{x\in\mathbb{R}} |H_{n,m}(x)| := \sup_{x\in\mathbb{R}} |(G_m(x) - G(x)) - (F_n(x) - F(x))| \longrightarrow 0 \quad \text{a.s.}, \quad n, m \to \infty.$$

Then by the continuous mapping theorem, we have

$$\sup_{t\in\mathbb{R}}\left|\int_0^t (G_m(x)-F_n(x))_+ - (G(x)-F(x))_+\mathrm{d}x\right| \leqslant \int_{\ell_F}^{\overline{z}} |H_{n,m}(x)|\mathrm{d}x \to 0 \quad \text{a.s.}, \tag{27}$$

where $\overline{z}$ is the right end point of the support of G. Similarly, we have

$$\sup_{t\in\mathbb{R}}\left|\int_{\ell_F}^t (F_n(x)-G_m(x))_+ - (F(x)-G(x))_+\mathrm{d}x\right| \to 0 \quad \text{a.s.} \tag{28}$$

Note that

$$\gamma_{\min} = \max_{t\in\mathbb{R}} \frac{\int_{\ell_F}^t (G(x)-F(x))_+\mathrm{d}x}{\int_{\ell_F}^t (F(x)-G(x))_+\mathrm{d}x} > 0. \tag{29}$$

Denote by $t_0 \in (\ell_G, \overline{z}]$ the maximizer of the optimization problem of (29). Then we have $\int_{\ell_F}^{t_0}(F(x)-G(x))_+\mathrm{d}x > 0$, and thus,

$$\tilde{\gamma}_{\min} \geqslant \frac{\int_{\ell_F}^{t_0} (G_m(x)-F_n(x))_+\,\mathrm{d}x}{\int_{\ell_F}^{t_0} (F_n(x)-G_m(x))_+\,\mathrm{d}x} \xrightarrow{\text{a.s.}} \frac{\int_{\ell_F}^{t_0} (G(x)-F(x))_+\mathrm{d}x}{\int_{\ell_F}^{t_0} (F(x)-G(x))_+\mathrm{d}x} = \gamma, \quad \text{as } n\to\infty. \tag{30}$$

To show the other direction, let $A(t) = \int_{\ell_F}^t (F(x)-G(x))_+\,\mathrm{d}x$, $B(t) = \int_{\ell_F}^t (G(x)-F(x))_+\,\mathrm{d}x$, $A_{n,m}(t) = \int_{\ell_F}^t (F_n(x)-G_m(x))_+\,\mathrm{d}x$ and $B_{n,m}(t) = \int_{\ell_F}^t (G_m(x)-F_n(x))_+\,\mathrm{d}x$. Note that for $t \in (\ell_F, \ell_G)$, $B(t) = 0$, $A(\ell_G) > 0$ and $A(t)$ and $B(t)$ are continuous increasing function. Then for any $\delta > 0$, there exists $\varepsilon \in (0, A(\ell_G))$ such that

$$\frac{B(t)+\varepsilon}{A(t)-\varepsilon} \leqslant \frac{B(t)}{A(t)} + \delta \quad \text{for} \quad t\in(\ell_G,\overline{z}), \quad \text{that is,} \quad \sup_{t\in(\ell_G,\overline{z})} \frac{B(t)+\varepsilon}{A(t)-\varepsilon} \leqslant \gamma+\delta.$$

By (27) and (28), there exists Ω' such that $\mathbb{P}(\Omega') = 1$ and $\sup_{t\in\mathbb{R}} |B_{n,m}(t) - B(t)| \to 0$ and $\sup_{t\in\mathbb{R}} |A_{n,m}(t) - A(t)| \to 0$ for any $\omega \in \Omega'$. From now on, we restrict on the fixed ω. Then for all $t \in (\ell_G, \overline{z})$, there exist n_0, m_0 such that for any $n \geqslant n_0$, $m \geqslant m_0$,

$$A_{n,m}(t) > A(t) - \varepsilon, \quad B_{n,m}(t) \leqslant B(t) + \varepsilon.$$

This implies that

$$\frac{B_{n,m}(t)}{A_{n,m}(t)} \leqslant \frac{B(t)+\varepsilon}{A(t)-\varepsilon}, \quad t\in(\ell_G, t_0).$$

and thus,

$$\sup_{t\in\mathbb{R}} \frac{B_{n,m}(t)}{A_{n,m}(t)} = \sup_{t\in(\ell_G,\overline{z})} \frac{B_{n,m}(t)}{A_{n,m}(t)} \leqslant \sup_{t\in(\ell_G,\overline{z})} \frac{B(t)+\varepsilon}{A(t)-\varepsilon} \leqslant \gamma+\delta,$$

where the first equality follows from $B_{n,m}(t) = 0$ a.s. for $t \in (\ell_F, \ell_G)$. This combined with (30) implies that $\tilde{\gamma}_{\min}$ is a consistent estimator of γ. □

A.5 Proof of Proposition 10 in Section 7

Proof of Proposition 10. By definition of D_h, we have $D_h(w+\alpha X - c(\alpha)) = w + \alpha D_h(X) - c(\alpha) =: D(X,\alpha)$. Since c is strictly convex, we have $\alpha^*(X_i) \in [0,\alpha_0]$ exists and is unique for $i = 1,2$. We show $\alpha^*(X_1) \leqslant \alpha^*(X_2)$ by contradiction. Suppose $\alpha^*(X_1) > \alpha^*(X_2)$. By Theorem 1, $P_h \geqslant \gamma$ implies that $\mathrm{DT}(h)$ is monotone in $\leqslant_\gamma^{\mathrm{S}}$, and thus, $D_h(X_1) \leqslant D_h(X_2)$. It follows that

$$\begin{aligned}
&D(X_1,\alpha^*(X_1)) - D(X_1,\alpha^*(X_2)) + D(X_2,\alpha^*(X_2)) - D(X_2,\alpha^*(X_1)) \\
&= \alpha^*(X_1)D_h(X_1) - \alpha^*(X_2)D_h(X_1) + \alpha^*(X_2)D_h(X_2) - \alpha^*(X_1)D_h(X_2) \\
&= (\alpha^*(X_1) - \alpha^*(X_2))(D_h(X_1) - D_h(X_2)) \leqslant 0,
\end{aligned}$$

where the inequality follows from $\alpha^*(X_1) > \alpha^*(X_2)$ and $D_h(X_1) \leqslant D_h(X_2)$. This yields a contradiction to that $\alpha^*(X_i)$ is the unique solution to $\max_{\alpha\in[0,\alpha_0]} D(X_i,\alpha)$, $i = 1,2$. Thus, we have $\alpha^*(X_1) \leqslant \alpha^*(X_2)$. □

B Simulation results for Section 5

We present some simulation results to complement the estimators in Section 5. Let F and G be cdfs of $N(\mu_F,\sigma_F^2)$ and $N(\mu_G,\sigma_G^2)$, respectively. Denote $\Delta\mu = \mu_G - \mu_F$ and $\Delta\sigma = \sigma_F - \sigma_G$. Müller et al. (2017) showed that $F \leqslant_\gamma^{\mathrm{S}} G$ with

$$\gamma \geqslant \frac{\int_{\frac{\Delta\mu}{\Delta\sigma}}^{\infty}(1-H(z))\mathrm{d}z}{\int_{\frac{\Delta\mu}{\Delta\sigma}}^{\infty}(1-H(z))\mathrm{d}z + \frac{\Delta\mu}{\Delta\sigma}} =: \gamma_{\min},$$

and by Theorem 2 (i) we can verify that $F \leqslant_c^{\mathrm{A}} G$ where $c \geqslant c_{\min} := 1/(1+\lambda_{\max})$ and

$$\lambda_{\max} = \frac{2(\mu_G - \mu_F)}{\sigma_F^2 - \sigma_G^2} = \frac{\Delta\mu}{\Delta\sigma}\frac{2}{\sigma_F + \sigma_G}.$$

In Table 2, we present the estimation of $\gamma_{\min}$ and $c_{\min}$ by simulations for normal distributions. We run $N = 1000$ replications, and for each time we set sample sizes to $n = 200$ and $n = 500$. From the results on Table 2, we can find that the mean squared error (MSE) of the estimate of $c_{\min}$ is smaller than that of $\gamma_{\min}$. The reason behind this phenomenon needs future research.

	F	G	$\gamma_{\min}$	$c_{\min}$	$\hat{\gamma}_{\min}$ (MSE)	$\hat{c}_{\min}$ (MSE)
$n = 200$	$N(0, 0.530^2)$	$N(0.013, 0.227^2)$	0.9	0.9	0.887 (0.0123)	0.892 (0.0105)
$n = 500$	$N(0, 0.530^2)$	$N(0.013, 0.227^2)$	0.9	0.9	0.894 (0.0069)	0.895 (0.0061)
$n = 200$	$N(0, 0.465^2)$	$N(0.038, 0.200^2)$	0.7	0.7	0.708 (0.0154)	0.700 (0.0112)
$n = 500$	$N(0, 0.465^2)$	$N(0.038, 0.200^2)$	0.7	0.7	0.706 (0.0063)	0.707 (0.0047)
$n = 200$	$N(0, 0.303^2)$	$N(0.015, 0.248^2)$	0.5	0.5	0.500 (0.0095)	0.503 (0.0055)
$n = 500$	$N(0, 0.303^2)$	$N(0.015, 0.248^2)$	0.5	0.5	0.502 (0.0033)	0.502 (0.0019)
$n = 200$	$N(0, 0.331^2)$	$N(0.03, 0.221^2)$	0.5	0.5	0.509 (0.0093)	0.508 (0.0058)
$n = 500$	$N(0, 0.331^2)$	$N(0.03, 0.221^2)$	0.5	0.5	0.504 (0.0034)	0.503 (0.0020)
$n = 200$	$N(0, 0.386^2)$	$N(0.06, 0.166^2)$	0.5	0.5	0.509 (0.0082)	0.508 (0.0051)
$n = 500$	$N(0, 0.386^2)$	$N(0.06, 0.166^2)$	0.5	0.5	0.503 (0.0031)	0.502 (0.0018)
$n = 200$	$N(0, 0.497^2)$	$N(0.12, 0.055^2)$	0.5	0.5	0.505 (0.0095)	0.505 (0.0057)
$n = 500$	$N(0, 0.497^2)$	$N(0.12, 0.055^2)$	0.5	0.5	0.501 (0.0034)	0.508 (0.0020)
$n = 200$	$N(0, 0.287^2)$	$N(0.078, 0.123^2)$	0.3	0.3	0.303 (0.0036)	0.301 (0.0017)
$n = 500$	$N(0, 0.287^2)$	$N(0.078, 0.123^2)$	0.3	0.3	0.300 (0.0013)	0.300 (0.0006)
$n = 200$	$N(0, 0.140^2)$	$N(0.072, 0.060^2)$	0.1	0.1	0.103 (0.0005)	0.099 (0.0002)
$n = 500$	$N(0, 0.140^2)$	$N(0.072, 0.060^2)$	0.1	0.1	0.101 (0.0003)	0.099 (0.0000)

Table 2: Simulated results of $\gamma_{\min}$ and $c_{\min}$ for normal distributions.