Selection and stochastic order

Thomas Noe
Balliol College/SBS*

1 October, 2016

Abstract

This paper derives necessary and sufficient conditions for stochastic and monotone likelihood ratio dominance conditioned on the selection of one random prospect over another when that prospect's realized value is higher. One of these conditions, geometric dominance, imposes simple, intuitive restrictions on compared distributions analogous to the monotone likelihood ordering. Using geometric dominance, we determine when selection preserves and reverses the unconditional stochastic order of random prospects. We then illustrate the power and applicability of selection-conditioned stochastic orders for resolving economic problems that hinge on selection through a series of illustrative applications. These applications examine the effect of admission contest structure on university-admission selection bias, the effect of the distributional properties of error laws on average and selection-conditioned treatment effects, and the conditions under which auction model bid predictions can be extended to predictions about winning bids.

Key Words: convexity, selection bias, stochastic dominance

The author would like to thank Alexei Boulatov and the participants in the Higher School of Economics research symposium for many helpful comments on an early draft of this paper. In addition, the author is grateful to Charles Manski, Todd Kaplan, and Thomas Hellmann, Jeffrey Ely, for many insightful comments on more recent drafts. The usual disclaimer applies.

*Saïd Business School, Park End Street, Oxford OX11HP UK, email:thomas.noe@sbs.ox.ac.uk

1 Introduction

The specter of selection bias haunts economics. Because much of economics is founded on rational choice, *competitive selection*—selection based on the optimizing decisions of agents—is arguably the preeminent selection problem. The aim of this paper is not to remove, correct, or test for competitive selection's effect, but rather to understand it, i.e., to understand the properties of random outcomes that determine the qualitative effect of competitive selection on inference.

Because the outcomes of choices are formalized as random variables and optimizing choice as choosing the best outcome, the act of competitive selection can be viewed as a choice between random variables, say $\tilde{X}$ and $\tilde{Y}$. Selection results in selection-conditioned distributions for $\tilde{X}$ and $\tilde{Y}$, represented respectively by $[\tilde{X}|\tilde{X} \geq \tilde{Y}]$ and $[\tilde{Y}|\tilde{Y} \geq \tilde{X}]$. Under what conditions can we infer from the dominance of $\tilde{X}$ over $\tilde{Y}$ from the dominance of $[\tilde{X}|\tilde{X} \geq \tilde{Y}]$ over $[\tilde{Y}|\tilde{Y} \geq \tilde{X}]$ and vice versa?

For many economic problems, the answers to these questions have significant implications, e.g., if a theoretical model that predicts more efficient corporate raiders make more generous takeover offers, does it also predict that more efficient raiders pay more for their acquisitions (i.e., submit higher selected winning bids)? Does the better performance of a drug on a self-selected sample of participants imply that it would have performed better if administered to all? Does the fact that one group of workers is less likely to be hired yet perform better if hired imply discrimination?[1]

Our analysis provides a complete answer to these sorts of questions through developing selection-conditioned versions of economic theory's two most pervasive size-based measures of dominance—(first-order) stochastic and monotone likelihood ratio (MLR) dominance (e.g., Hanoch and Levy, 1969; Milgrom, 1981). Like stochastic dominance and MLR dominance, these selection-conditioned measures impose restrictions on the marginal unconditional distributions of compared outcomes which are necessary and sufficient for one outcome to dominate another. In contrast to unconditional dominance orders, these conditions ensure dominance relations between the selection-conditioned outcomes rather than the unconditional outcomes. We term the condition for selection-conditioned stochastic dominance, *competitive selection stochastic dominance* (CSSD); we term the condition for selection-conditioned monotone likelihood ratio (MLR) dominance, *geometric dominance*.

The selection-conditioned analog of stochastic dominance, CSSD, is not transitive and thus the CSSD relation does not define an order relation between distributions. Moreover, the condition for CSSD dominance is far more complex than the condition imposed by stochastic

[1] As will become apparent, the only sort of selection conditioning considered in this paper is competitive selection. Thus, in the sequel, we will refer to "competitive selection" simply as "selection" unless, in the discussion of a particular result, we are comparing competitive selection with some other type of selection.

dominance. In short, this paper demonstrates that there is no ordering of selection-conditioned random outcomes based on their unconditional distributions which is analogous to the stochastic dominance ordering of unconditional outcomes.

In contrast, the condition for the selection-conditioned analog of MLR dominance, geometric dominance, is simple and defines an order relation. While MLR ordering requires the quantile transform relating the compared distributions to be convex, geometric dominance requires that the quantile transformation be *geometrically convex*, roughly speaking, convex if plotted on a log-log scaled grid.

When distributions are absolutely continuous, geometric dominance is defined by a simple condition—the ratio between the dominant and dominated distributions' logarithmic derivatives (i.e., their reversed hazard rates) must be increasing. This condition is reminiscent of the MLR dominance condition—the ratio between the dominant and dominated distributions' derivatives (their PDFs) being increasing. Moreover, just as MLR ordering is a sufficient condition for stochastic dominance, geometric dominance is a sufficient condition for CSSD dominance, stochastic dominance conditioned on selection.[2] In short, this paper shows that there is an ordering of selection-conditioned random prospects analogous to the MLR ordering of unconditional prospects, an ordering we term *geometric dominance*.

As well as being analogous to MLR ordering, geometric dominance has an intuitive interpretation. Geometric dominance is equivalent to increasing quantile elasticity of the relative performance of the geometrically dominant distribution, i.e., the marginal relative performance of the dominant distribution increases *proportionally* with the quantile at which the geometrically dominant and dominated distributions are compared. Thus, a distribution may well be inferior at every quantile (e.g., MLR dominated), but if its relative inferiority is falling vary steeply as the comparison quantile increases, the unconditionally inferior distribution can be geometrically dominant and thus MLR dominant conditioned on selection. However, we show that, for almost all size-indexed families of textbook statistical distributions, the geometric dominance ordering is consistent with unconditional stochastic dominance ordering. Thus, for such families, selection preserves inference.

Having completed the theoretical development, we turn to applications. The first two applications consider a specific policy-relevant economic problem: does a higher selection probability for applicants from group A relative to group B, combined with lower ex post performance of selected applicants from group A, imply biased, non-performance based, selection? The answer to this question has significant implications in empirical contexts in which both ex ante selection probabilities and ex post performance can be objectively measured, e.g., mortgage

[2]Unfortunately as well as sharing many of the advantages of MLR ordering, geometric dominance also shares its limitations. Like MLR ordering, geometric dominance ordering characterizes only marginal distributions, and thus cannot by its very nature, determine the effect of joint distribution structures, correlation, or copulas, on selection dominance.

lending to minority applicants (e.g., Han, 2004), discrimination in professional sports, (Lavoie et al., 1987), and university admissions. Although the analysis applies equally well to all of these contexts, for the sake of definiteness, we interpret the model using the university admissions frame.

Thus, the first application models university admissions competition as a classic Lazear and Rosen (1981) tournament. Verifying the geometric dominance of the high-ability student group shows that selection always preserves inference: When students from the two groups compete for places, under the assumption of non-discriminatory quality-based selection, the post-admission performance of admitted students from one group will be better than another group's performance if and only if the unconditional quality of applicants from that group is higher. Because the unconditional quality of applicants from the more able group is higher, members of the more able group will also be more likely to be selected. Thus, the "Oxford paradox"—state-school applicants are less likely to be accepted by Oxford, yet, if accepted, perform better—is indeed somewhat paradoxical in the tournament framework.

The second application, developed under the same parametric assumptions as the first, models university admissions competition in the effort-bidding framework developed in Baye et al. (1993, 1996). In this setting, verifying the geometric dominance of the low-ability student group shows selection always *reverses* inference: members of the higher quality group will always be more likely to be admitted yet always perform worse on average post admission. Thus, the Oxford paradox is not very paradoxical in the effort-bidding framework.[3] Both the tournament model and effort-bidding model yield very similar predictions about the relation between model parameters and the unconditional distribution of performance, but contrary predictions about selection-conditioned performance.

The third application considers the problem of identifying the sign of average treatment effects from the sign of selection-conditioned treatment effect (e.g., Manski, 1990; Heckman, 1979; Heckman and Honoré, 1990). Results developed in this paper provide a condition on the error distribution, satisfied by textbook error laws, which ensures identification of the sign of average treatment effects from selection-conditioned treatment effects. Next, we show that the elasticity characterization of geometric dominance provides a "recipe" for constructing counterexamples to selection preservation. A recipe which generates a simple symmetric, unimodal, non fat-tailed error law, common to both treatments, under which geometric dominance, and thus identification, fails.

Finally, in our last application, we apply our results to first-price auctions to determine the relation between the distribution of bids and the distribution of winning bids. Identifying the winning-bid distribution is quite important in auctions. First, as Laffont et al. (1995) point out,

[3] For Oxford admission statistics see Gazette (2016). For a discussion of achievement and qualification of state and independent school students in the UK, see Garner (2015) and Machin et al. (2009). For a more general discussion of selection bias and the evaluation of university admissions, see Stevens et al. (2016).

frequently bid data is only available for the winning bid because only the winning bid is paid. Second, in first-price and all-pay auctions, winning bids determine the terms of exchange. We show that in the Maskin and Riley (2000) model of asymmetric first-price auctions, geometric dominance of the strong bidder's bid distribution is equivalent to the elasticity of bid-shading being greater for the strong bidder. Not only is this result immediate from our earlier analysis, it is also quite intuitive given the elasticity characterization of geometric convexity. We then show that the bid elasticity condition is satisfied in the parametric asymmetric auction model of Kaplan and Zamir (2012). Thus, geometric convexity can be used to infer from the fact that strong bidders bid more, that strong bidders pay more.

Note, that, in order to obtain these results, it was not necessary impose any additional assumptions on the models considered. Microeconomic models typically impose restrictions on stochastic shocks, e.g., parametric forms, MLR ordering, logconcavity of the probability density functions, and especially in the auction literature, restrictions on hazard rates and reversed hazard rates (e.g., McAfee and McMillan, 1987; Kirkegaard, 2012). In these examples, and we conjecture in many other cases, these restrictions imply geometric dominance ordering. Thus, the theorist who has "bought" a model solution to her model with such restrictions, frequently also has, unknowingly perhaps, bought selection robust predictions as well.

In general, the paper is part of a vast research program devoted to using stochastic orders to sign and comparative static relations.(Quah and Strulovici, 2009; Hanoch and Levy, 1969; Athey, 2002; Milgrom, 1981, *inter alia*). The problem considered in this paper is also closely related to research on identifying the sign of treatment effects in the presence of selection bias (e.g., Manski, 1990; Heckman, 1979; Heckman and Honoré, 1990). However, our perspective is very different. Rather than aiming to combat selection bias, we simply aim to develop an understanding of the characteristics random errors that generate it. Finally, the technical development is indebted to recent mathematical research related to supermultiplicative functions and geometric convexity, (e.g., Finol and Wójtowicz, 2000; Niculescu and Persson, 2004).

2 Selection and inference: An example

The following example both shows that the question we consider—the relation between conditional and unconditional dominance—is not trivial and also suggests the sort of analysis required to answer it.

Example 1. Consider a random variable, $\tilde{Y} \overset{d}{\sim} G$, where G is the uniform distribution over $(0,1)$. Let $\tilde{X} \overset{d}{\sim} F$, where $F(x) = p + (1-p)\,G(x), \quad p \in (0,1),\, x \in [0,1]$.[4] The distribution of $\tilde{X}$ is a convex combination of a point mass at 0 and the distribution of $\tilde{Y}$. Let $\tilde{Y}'$ be a random variable

[4]For all distribution functions considered in Example 1, the distribution function evaluated at $x < 0$ is defined to equal 0 and, evaluated at $x > 1$ is defined to equal 1.

independent of $\tilde{Y}$, which is also uniformly distributed between 0 and 1. We can think of $\tilde{X}$ as being a gamble between 0 and $\tilde{Y}'$, with the gamble leading to 0 with probability p and leading to $\tilde{Y}'$ with probability $1-p$.[5] $\tilde{X}$ is selected, i.e., $\tilde{X} > \tilde{Y}$, if and only if (a) the gamble resulted in $\tilde{Y}'$ and (b) $\tilde{Y}' > \tilde{Y}$. In which case, $\tilde{X} = \tilde{Y}' = \max[\tilde{Y}, \tilde{Y}']$. Next note that $\mathbb{P}[\max[\tilde{Y}, \tilde{Y}'] < x]$ is the distribution of the maximum of the two independent draws from a uniform distribution over [0,1]. Thus, $\mathbb{P}[\max[\tilde{Y}, \tilde{Y}'] < x] = x^2$, $x \in [0,1]$. Hence, the selection-conditioned distribution of $\tilde{X}$, represented by $F^c(x) = \mathbb{P}[\tilde{X} \le x | (\tilde{X} > \tilde{Y}]$, is given by $F^c(x) = x^2$, $x \in [0,1]$.

Now consider the selection-conditioned distribution of $\tilde{Y}$. $\tilde{Y}$ is selected if and only if either (a) $\tilde{X} = 0$ or (b) $\tilde{X} = \tilde{Y}'$. In case (a) $\tilde{Y}$ is always selected, so the probability that $\{\tilde{Y} \le x\}$ given that $\tilde{X} = 0$ and $\tilde{Y}$ is selected is simply $\mathbb{P}[\tilde{Y} \le x]$. Thus in case (a), the selection-conditioned probability of distribution of $\tilde{Y}$ is simply its unconditional distribution, $G(x) = x$, $x \in [0,1]$. In case (b), $\tilde{Y}$ is selected if and only if $\tilde{Y} > \tilde{Y}'$, i.e., $\tilde{Y} = \max[\tilde{Y}, \tilde{Y}']$. Thus, in case (b), the selection-conditioned probability of distribution of $\tilde{Y}$ is given by $\mathbb{P}[\max[\tilde{Y}, \tilde{Y}'] < x] = x^2$, $x \in [0,1]$.

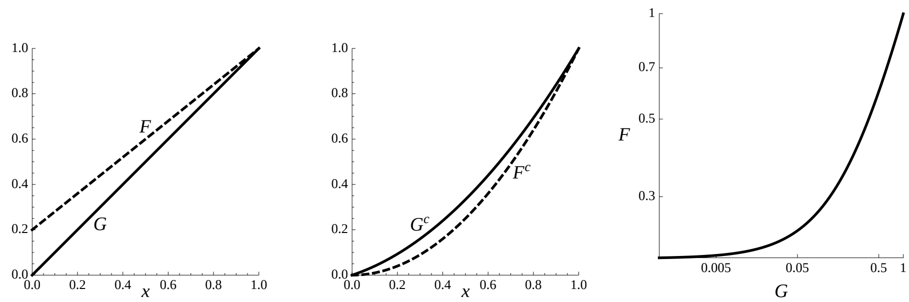
The probability of $\tilde{Y}$ being selected, $\mathbb{P}[\tilde{Y} > \tilde{X}]$, is 1 if $\tilde{X} = 0$ and $1/2$ if $\tilde{X} > 0$. Thus, the overall probability that $\tilde{Y}$ is selected is $p + 1/2\,(1-p)$. Hence, the probabilities of case (a) and case (b) are given by

$$\text{(a): } \mathbb{P}\big[\tilde{X} = 0 \big| \tilde{Y} > \tilde{X}\big] = \frac{p}{p + 1/2\,(1-p)}, \text{ (b): } \mathbb{P}\big[\tilde{X} > 0 \big| \tilde{Y} > \tilde{X}\big] = \frac{1/2\,(1-p)}{p + 1/2\,(1-p)}. \tag{1}$$

The selection-conditioned distribution of $\tilde{Y}$ is the probability-weighted average of the selection-conditioned distributions in the two cases, i.e., G^c, the selection-conditioned distribution of $\tilde{Y}$, is given by

$$G^c(x) = \mathbb{P}\big[\tilde{Y} \le x \big| \tilde{Y} > \tilde{X}\big] = \frac{p}{p + \frac{1}{2}(1-p)} x + \frac{\frac{1}{2}(1-p)}{p + \frac{1}{2}(1-p)} x^2, \quad x \in [0,1].$$

The unconditional and selection-conditioned distributions of $\tilde{X}$ and $\tilde{Y}$ are depicted in Figure 1.



A. Distributions of F and G B. Selection-conditioned distributions of F and G C. Log-log scaled *P*-*P* plot of F against G

Figure 1: *Example of the effect of selection.* The figures depict the effect of selection on the two compared random variables in Example 1. In the figure, $p = 0.20$.

[5] The derivation of the conditional distributions provided below is somewhat informal. For a more formal, albeit rather mechanical, derivation, see Online Supplement S4.

In this example 1, $\tilde{X}$ is stochastically dominated by $\tilde{Y}$ (Figure 1.A). Yet conditioned on selection $\tilde{X}$ stochastically dominates $\tilde{Y}$ (Figure 1.B). So selection reverses inference, implying that naive attempts to infer the unconditional "quality" of $\tilde{X}$ and $\tilde{Y}$ by comparing their selection-conditioned quality will always result in incorrect inferences. The basis for selection reversing dominance is the effect of zero draws from $\tilde{X}$. Whenever a zero draw of $\tilde{X}$ is compared with a draw from $\tilde{Y}$, the draw from $\tilde{Y}$ is selected regardless of its quality. This effect, which one might term an "admission effect" permits low realizations of $\tilde{Y}$ to be "admitted" into the selected sample. Moreover, because $\tilde{Y}$ places all its mass above zero, zero draws from $\tilde{X}$ are never selected and thus these draws have no effect on the quality of selected draws from $\tilde{X}$.

More generally, the distinguishing feature of the relation between F and G is that the relative unconditional inferiority of F is decreasing rapidly in the quantile of G at which the distributions are compared: evaluated at $x = G(x) = 0.05$, a 10% increase in x increases G by 10% and F by $1.6\bar{6}$%; evaluated at $x = G(x) = 0.85$, a 10% increase in x increases G by 10% and increases F by approximately $7.6\bar{6}$%. The increasing proportional response of F to proportional increases in G is illustrated in Panel C of Figure 1, a Log–log *P-P* plot, i.e., a plot F against G on a log-scaled grid.

If we applied the same increasing transformation to F and G, the conditional and unconditional distributions of F and G might well be transformed beyond recognition. However, selection reversal would not be affected nor would the graph of the relation between F with G depicted in Figure 1.C. Thus, what seems to matter for selection reversal is the proportionality relation between the *quantiles* of the compared distributions rather the "shape" properties of either distribution considered in isolation from the other.

This paper formalizes and generalizes these observations to provide conditions on the unconditional distributions of compared prospects that are necessary and sufficient to rule in and rule out cases like Example 1 and thus to determine the qualitative effect of selection on inference. Generalizing requires a tool for relating the quantiles of the compared distributions. Fortunately, this tool, the *quantile transformation function*, has already been developed in the economic literature on stochastic dominance. Generalizing also requires assimilating the quantile elasticity intuition into a precise mathematical concept. Again, fortunately, this concept has been developed, albeit for very different purposes, in the mathematics literature—*geometric convexity*. Finally, the results thus formalized need to be shown to be applicable to understanding the wide variety of interesting economic problems that hinge on selection.

3 Basic results

3.1 The problem

Our aim is to compare distribution functions and determine the properties of these distribution functions which ensure that conditioned on selection, i.e., conditioned on draws from each distribution being selected only if they exceed the draws from the other, one distribution stochastically dominates the other. To avoid the problem of ties and indeterminacies, we impose the following restrictions on the distribution functions we consider:

Definition 1. Distribution functions F and G form an *admissible pair of distributions* if

(i) F and G have common support $[\underline{x}, \bar{x}]$, $-\infty \leq \underline{x} < \bar{x} \leq \infty$.

(ii) G is continuous and absolutely continuous with respect to F and F is continuous except perhaps at $\underline{x}$.

A collection of distribution functions is admissible if all pairs in the collection are admissible.

Note that condition (i) does not rule out unbounded supports. It does rule out gaps in the support. This no-gaps condition is not necessary to derive our results. However, absent this condition, stating some of our results would become much more cumbersome. If we allowed for gaps in the common support of the compared distributions, we would need to identify all intervals of constancy of the two distribution functions as equivalent, and then state our results in terms of the resulting equivalence classes. The assumption that the supports of the two distributions are identical involves little loss of generality. First, suppose that the upper bound of support of F exceeds the upper bound of the support of G. As will become apparent in the sequel, F could never be selection dominated by G. If the lower bound of the support of F were less than the lower bound of the support of G, then draws from F below the lower bound of G's support would never be selected. Thus, the selection-conditioned distribution of F would be the same as it would have been if F simply had a point mass at the lower bound of G's support. This case is analyzed.

Condition (ii) eliminates the problem of ties. This assumption combined with the assumption that F is continuous except perhaps at $\underline{x}$ ensures that on their common support F and G are continuous and strictly increasing. Note that although we assume that F and G are continuous on their common support, we do not assume that F or G are absolutely continuous with respect to Lebesgue measure.

3.2 Competitive selection stochastic dominance (CSSD)

In the case of unconditional comparisons of distributions, the stochastic dominance partial order, $\underset{sd}{\succcurlyeq}$ is defined as follows: distribution F stochastically dominates distribution G if the expectation of all non-constant increasing functions under F is greater than or equal to the their expectation under G.[6] It is well known that the stochastic dominance condition is equivalent to

$$F \underset{sd}{\succcurlyeq} G \text{ if } F(x) \leq G(x), x \in (\underline{x}, \bar{x}),$$

If the inequality in expression (3.2) is strict we will say that F *strictly stochastically dominates* G. This condition is equivalent to the expectation of all non-constant increasing function being higher under F than under G.[7]

We seek an analogous definition stochastic dominance conditioned on selection. Suppose that $\tilde{X}$ and $\tilde{Y}$ are random variables whose distribution functions, F and G respectively, form an admissible pair. Let v be the valuation, i.e., an increasing function, attached to realizations of the random variables. Selection is competitive and based on the ranking of the two random variables. Thus, the expected value of X under v is conditioned on $\{\tilde{X} \geq \tilde{Y}\}$ The expected value of $\tilde{Y}$ under v is conditioned on $\{\tilde{Y} \geq \tilde{X}\}$. Our basic question is when will the expected valuation of $\tilde{X}$ conditioned on $\tilde{X}$ being selected exceed the expected valuation of $\tilde{Y}$ conditioned on $\tilde{Y}$ being selected for all increasing non-constant valuation functions? This question motivates the following definition:

Definition 2. We say that $\tilde{X}$ (or its distribution function F) dominates $\tilde{Y}$ (or its distribution function G) under *competitive selection stochastic dominance* (CSSD) if, for all integrable increasing, non constant functions $v : [\underline{x}, \bar{x}] \to \Re$,

$$\mathbb{E}[v(\tilde{X})|\tilde{X} > \tilde{Y}] \geq \mathbb{E}[v(\tilde{Y})|\tilde{Y} > \tilde{X}]. \tag{2}$$

If the inequality in (2) holds strictly, we will say that $\tilde{X}$ *strictly CSSD dominates* $\tilde{Y}$.

Note that the definition of admissible pairs ensures that $\tilde{X} \neq \tilde{Y}$ with probability 1, so, expression (2) in Definition 2, is equivalent to

$$\mathbb{E}[v(\tilde{X})|\tilde{X} \geq \tilde{Y}] \geq \mathbb{E}[v(\tilde{Y})|\tilde{Y} \geq \tilde{X}].$$

Like stochastic dominance, CSSD dominance is preserved if the compared distributions are independent and both are both shocked by a common factor. Thus, the CSSD relation can be extended to handle a very commonly assumed form of dependence between compared random variables—the signal-plus-noise formulation—in which the compared variables are the sum (or product) a common factor term and an idiosyncratic factor. In fact, far more general forms of dependency on the common shocks preserve CSSD as the following result verifies.

[6]In this paper "increasing" means that a function is weakly monotonically increasing (e.g., nondecreasing) and "strictly increasing" means that it is strongly monotonically increasing.

[7]Admittedly this is a slightly unorthodox use of the term "strict" stochastic dominance. Strict stochastic dominance is typically is defined as the stochastic dominance condition, expression (3.2), being satisfied always with strict inequality holding at some point. However, when we develop our most important ordering relation, geometric dominance, it will be defined based a convexity relation. The conventional definition of strict convexity requires that, on the interior of a function's support, the convexity inequality is always strictly satisfied not that the convexity inequality is weakly satisfied at all points and strictly at one. Rather than introduce yet more new terminology, e.g., "strong stochastic dominance," we choose to define strict stochastic dominance analogously with strict convexity.

Lemma 1. *Let, $\tilde{X} \overset{d}{\sim} F$, $\tilde{Y} \overset{d}{\sim} G$, $\tilde{Z} \overset{d}{\sim} H$ be random variables. Suppose that F and G form an admissible pair and that $\tilde{Z}$ is jointly independent of $\tilde{X}$ and $\tilde{Y}$, H is an absolutely continuous with probability density function, h, supported by S. Let $\Gamma : S \times]\underline{x}, \bar{x}] \to \Re$, $S \subset \Re$ be a function such that for all $z \in S$, $x \hookrightarrow \Gamma(z,x)$ is strictly increasing. Finally, suppose that expectations exist under Γ, i.e.,*

$$\max \left[\mathbb{E}\left[|\Gamma(\tilde{Z},\tilde{X})| \right], \mathbb{E}\left[|\Gamma(\tilde{Z},\tilde{Y})| \right] \right] < \infty.$$

Then, if $\tilde{X}$ CSSD dominates $\tilde{Y}$, $\Gamma(\tilde{Z},\tilde{X})$ CSSD dominates $\Gamma(\tilde{Z},\tilde{Y})$.

Proof. See Supplement S1.

In the subsequent theoretical development, *we assume that $\tilde{X}$ and $\tilde{Y}$ are independent*. This is a natural assumption because we are attempting to find analogs of unconditional stochastic ordering conditions which themselves are defined in terms of marginal distributions. We will however, in the applications section, Section 12, use Lemma 1 to apply our results to cases where the compared prospects are dependent.

3.3 Conditions for CSSD

Note that, for admissible pairs of distributions, we can express conditioning on competitive selection as follows:

$$\mathbb{E}[v(\tilde{X}) | \tilde{X} > \tilde{Y}] = \frac{\mathbb{E}[v(\tilde{X})\, \mathbf{I}_{\tilde{X}>\tilde{Y}}]}{\mathbb{E}[\mathbf{I}_{\tilde{X}>\tilde{Y}}]} \quad \text{and} \quad \mathbb{E}[v(\tilde{Y}) | \tilde{Y} > \tilde{X}] = \frac{\mathbb{E}[v(\tilde{Y})\, \mathbf{I}_{\tilde{Y}>\tilde{X}}]}{\mathbb{E}[\mathbf{I}_{\tilde{Y}>\tilde{X}}]}, \tag{3}$$

where, in expression (3), $\mathbf{I}$ represents the set indicator function of the subscripted set. By the independence of $\tilde{X}$ and $\tilde{Y}$ and Fubini's Theorem,

$$\begin{aligned} \mathbb{E}[v(\tilde{X})\, \mathbf{I}_{\tilde{X}>\tilde{Y}}] &= \int_{\underline{x}}^{\bar{x}} \int_{\underline{x}}^{\bar{x}} v(x)\, \mathbf{I}_{y<x}\, dF(x)\, dG(y) \\ &= \int_{\underline{x}}^{\bar{x}} v(x) \left(\int_{\underline{x}}^{\bar{x}} \mathbf{I}_{y<x}\, dG(y) \right) dF(x) = \int_{\underline{x}}^{\bar{x}} v(x)\, G(x)\, dF(x). \end{aligned} \tag{4}$$

Using the same reformulation, we can express the other components of the conditional expectations as follows:

$$\begin{gathered} \mathbb{E}[v(\tilde{Y})\, \mathbf{I}_{\tilde{Y}>\tilde{X}}] = \int_{\underline{x}}^{\bar{x}} v(x)\, F(x)\, dG(x), \\ \mathbb{E}[\mathbf{I}_{\tilde{X}>\tilde{Y}}] = \int_{\underline{x}}^{\bar{x}} G(z)\, dF(z), \quad \mathbb{E}[\mathbf{I}_{\tilde{Y}>\tilde{X}}] = \int_{\underline{x}}^{\bar{x}} F(z)\, dG(z). \end{gathered} \tag{5}$$

Using the expressions in (4) and (5), we can express the conditional expectations as follows:

$$\mathbb{E}[v(\tilde{X}) | \tilde{X} > \tilde{Y}] = \frac{\int_{\underline{x}}^{\bar{x}} v(z) G(z)\, dF(z)}{\int_{\underline{x}}^{\bar{x}} G(z)\, dF(z)} \quad \text{and} \quad \mathbb{E}[v(\tilde{Y}) | \tilde{Y} > \tilde{X}] = \frac{\int_{\underline{x}}^{\bar{x}} v(z) F(z)\, dG(z)}{\int_{\underline{x}}^{\bar{x}} F(z)\, dG(z)}.$$

Define the probability distribution functions H and J by

$$H(x) = \frac{\int_{\underline{x}}^{x} G(z)\, dF(z)}{\int_{\underline{x}}^{\bar{x}} G(z)\, dF(z)} \quad \text{and} \quad J(x) = \frac{\int_{\underline{x}}^{x} F(z)\, dG(z)}{\int_{\underline{x}}^{\bar{x}} F(z)\, dG(z)}. \tag{6}$$

Using H and J, we can express the conditioning relations as simple expectations over distributions:

$$\mathbb{E}[v(\tilde{X})|\tilde{X} > \tilde{Y}] = \int_{\underline{x}}^{\bar{x}} v(z)\, dH(z) \quad \text{and} \quad \mathbb{E}[v(\tilde{Y})|\tilde{Y} > \tilde{X}] = \int_{\underline{x}}^{\bar{x}} v(z)\, dJ(z).$$

Thus, for competitive selection stochastic dominance (CSSD), formalized in Definition 2, to hold, it is necessary and sufficient that H stochastically dominate J, i.e., $H \le J$. From equation (6), we see that stochastic dominance of H over J is equivalent to the condition that, for all $x \in (\underline{x}, \bar{x})$,

$$\frac{\int_{\underline{x}}^{x} G(z)\, dF(z)}{\int_{\underline{x}}^{x} F(z)\, dG(z)} \le \frac{\int_{\underline{x}}^{\bar{x}} G(z)\, dF(z)}{\int_{\underline{x}}^{\bar{x}} F(z)\, dG(z)}. \tag{7}$$

Integration by parts yields the following equivalent condition for (7).

$$\frac{1}{F(x)\, G(x)} \int_{\underline{x}}^{x} F(z)\, dG(z) \ge \int_{\underline{x}}^{\bar{x}} F(z)\, dG(z). \tag{8}$$

3.4 Competitive selection MLR dominance

First note that for two absolutely continuous distributions, F, and G with common support $[\underline{x}, \bar{x}]$ with densities f and g respectively, F dominates G in the MLR order if the ratio f/g is increasing over $(\underline{x}, \bar{x})$ and strictly dominates G if this ratio is strictly increasing.

We seek an analog of this condition for selection-conditioned distributions. The analog is competitive selection MLR dominance.

Definition 3. If F and G are an admissible pair of absolutely continuous distributions, F dominates G in the *competitive selection MLR ordering*, if the selection-conditioned distribution of F dominates the selection-conditioned distribution of G under the MLR order.[8]

In contrast to CSSD, the distributional restrictions imposed by the selection-conditioned analog of MLR ordering are simple and recorded in the next Lemma.

Lemma 2. *If F and G are absolutely continuous with probability density functions, f and g respectively and $x \hookrightarrow (f(x)\, G(x))/(g(x)\, F(x))$ is increasing then F dominates G in the* competitive selection conditioned MLR *ordering and strictly dominates if $x \hookrightarrow (f(x)\, G(x))/(g(x)\, F(x))$ is strictly increasing.*

[8]In contrast to stochastic dominance, note that we define MLR, and thus competitive selection-conditioned MLR, in terms of restrictions on distribution functions rather than in terms of the ordering it imposes on expected valuation for a class of valuation functions. Our approach is "inconsistent" only because the orderings themselves are "inconsistent." Unlike stochastic dominance, MLR cannot be defined based on expected valuations (Lehmann and Rojo, 1992). Also note that it is possible to define MLR more generally in terms of the Radon-Nykodim derivative of the compared distributions. However, since this approach is both unfamiliar to many readers and the extra generality has no application in this paper, we choose to use the more conventional definition which restricts attention to absolutely continuous distributions.

Proof. The densities of the selection-conditioned distributions, H and J, defined in equation (7), are

$$h(x) = f(x)G(x), \quad j(x) = g(x)F(x), \quad x \in (\underline{x}, \bar{x}).$$

Thus, if the ratio h/j is increasing, the selection-conditioned distribution of F, H dominates the selection-conditioned distribution of G, J, in the MLR ordering. □

Using the well-known properties of the MLR ordering, we see that competitive selection conditioned MLR imposes even stronger restrictions on the expected selection-conditioned valuations than CSSD.

Lemma 3. *Let v be an increasing integrable function. Let B be a measurable set, and assume that $\mathbb{P}[\tilde{X} \in B] > 0$ and $\mathbb{P}[\tilde{Y} \in B] > 0$. Let $\tilde{X}$ and $\tilde{Y}$ be independent random variables with distribution functions F and G respectively. Then if F competitive selection conditioned MLR dominates G then*

$$\mathbb{E}[v(\tilde{X}) | \tilde{X} > \tilde{Y} \;\&\; \tilde{X} \in B] \geq \mathbb{E}[v(\tilde{Y}) | \tilde{Y} > \tilde{X} \;\&\; \tilde{Y} \in B],$$

with the weak inequality being replaced by a strict inequality if the dominance is strict.

Proof. The result follows directly Lemma 2 and Theorem 1.C.2 in Shaked and Shanthikumar (1994). □

Thus, if two distributions are ordered by competitive selection-conditioned MLR, the dominant distribution not only produces higher expected valuation conditioned on selection but also when conditioning is with respect to both selection and any measurable subset of the range of possible realizations.

4 Probability distributions, the u transform, and CSSD

4.1 The u transform

Although expression (8) does provide a condition for CSSD, the import of this condition is rather opaque. To elucidate the restrictions imposed by inequality (8), we follow a standard approach in microeconomics and statistics for modeling stochastic dominance and MLR dominance relations: we develop the CSSD condition in terms of the quantile transform function, $u = F \circ G^{-1}$ (where $\circ$ represents functional composition).[9] The basic motivation using the quantile transform to model stochastic order relations derives from noting that only the behavior of the distributions relative to each other matters for dominance.

[9] For an example of the application of the quantile transform approach in microeconomics, see MIT open courseware http://ocw.mit.edu/courses/economics/14-123-microeconomic-theory-iii-spring-2015/lecture-notes-and-slides/MIT14MIT14_123S15_Chapt4.pdf. For an example from statistics, see Lehmann and Rojo (1992)

First note that, because regular pairs of distributions are continuous and strictly increase over their support, G and F have well-defined strictly increasing inverse functions. We represent the inverse function of F with F^{-1} and the inverse function of G with G^{-1}. Thus, the function $u = F \circ G^{-1}$ is well defined and $F = u \circ G$, i.e., u transforms G into F. We will refer to u as the *quantile transform function* or simply the *transform function*. If $\tilde{X} \overset{d}{\sim} F$ and $\tilde{Y} \overset{d}{\sim} G$ then, $u(t) = F(G^{-1}(t))$ represents the probability that $\tilde{X}$ is less than the t-th quantile of the distribution of Y, $G^{-1}(t)$. Thus u can be thought of as receiving quantiles from G and "pushing" them onto the corresponding quantiles of F.

Standard stochastic orders can be expressed quite simply using the transform function. Stochastic dominance is equivalent to $u(t) \leq t$. To see this note that, $u(t) = F \circ G^{-1}(t) \leq t$ if and only if $F(G^{-1}(G(x))) \leq G(x)$, i.e., $F(x) \leq G(x)$. Intuitively, when F stochastically dominates G, u pushes quantiles of G into lower quantiles of F.

The fact that G and F regularly related implies that u is continuous on $(0,1)$ and that $\lim_{t\uparrow 1} u(t) = 1$ and $\lim_{t\downarrow 0} u(t) = F(\underline{x})$ Thus, we can extend the definition of the quantile transform function, u, by defining $u(1) = 1$ and $u(0) = F(\underline{x})$. The resulting function, $u : [0,1] \to [0,1]$, will be strictly increasing and continuous.

Definition 4. If a function $u\colon [0,1] \to [0,1]$ is strictly increasing and continuous, with $u(0) < 1$ and $u(1) = 1$, we will call u an *admissible function*.

4.2 Conditions for CSSD dominance

These observations provide the tools required to derive the restriction on the relation between quantiles of the compared distributions implied by selection-conditioned stochastic dominance. This restriction is recorded below.

Theorem 1. *Suppose that F and G are an admissible pair of distribution functions; Let $u = F \circ G^{-1}$, then F (strictly) CSSD dominates G, if and only if the quantile transform function, $u = F \circ G^{-1}$, satisfies condition that*

$$\text{for all } t \in (0,1], \quad \frac{1}{tu(t)} \int_0^t u(s)\, ds \geq \int_0^1 u(s)\, ds.$$

Proof. Using the u transform, we can express equation (8) as

$$\frac{1}{u \circ G(x)\, G(x)} \int_{\underline{x}}^{x} u \circ G(z)\, dG(z) \geq \int_{\underline{x}}^{\bar{x}} u \circ G(z)\, dG(z). \tag{9}$$

Using the change of variables $t = G(x)$ on CSSD condition given by equation (9) shows that the expression in the theorem is equivalent to equation (9). □

The CSSD relation is actually a relation between the transform function and its average. To see this, define for any admissible u, the average transform function, U, as follows:

$$U(t) = \begin{cases} \frac{1}{t}\int_0^t u(s)\,ds & \text{if } t \in (0,1], \\ u(0) & \text{if } t = 0. \end{cases}. \tag{10}$$

$U(t)$ represents the average value of u over the interval $[0,t]$. Note that U is continuous over $[0,1]$ and that the continuity of u ensures that U is continuously differentiable over $(0,1]$. Next, note that $(U(t)\,t)' = u(t)$. Therefore,

$$t\,U'(t) + U(t) = u(t), \quad t \in (0,1]. \tag{11}$$

Expressed using the average transform function, CSSD dominance of Theorem 1 is thus equivalent to

$$\text{for all } t \in (0,1],\ \frac{U(t)}{u(t)} \geq U(1). \tag{12}$$

Although Theorem 1 provides necessary and sufficient conditions for CSSD, the relation between the unconditional distributions which provides the necessary and sufficient conditions is opaque. For this reason, we develop two sufficient conditions for CSSD. These conditions are more transparently related to the underlying unconditional distributions. The first and perhaps most obvious sufficient condition for the satisfaction of the CSSD condition provided by(12) is recorded in Lemma 4 below.

Lemma 4. *If*

$$t \hookrightarrow \frac{u(t)}{U(t)} \text{ is (strictly) increasing then } F \text{ (strictly) CSSD dominates } G.$$

Proof. $t \hookrightarrow u(t)/U(t)$ being increasing implies that $t \hookrightarrow U(t)/u(t)$ is decreasing and thus, the necessary and sufficient condition for CSSD, given by expression (12) is satisfied. □

To develop the second sufficient condition, first note that, by a change of variables in expression (1), we can rewrite the CSSD condition as follows

$$\text{For all } t \in (0,1),\ \int_0^1 \Big(u(t\,s) - u(s)\,u(t)\Big)\,ds \geq 0. \tag{13}$$

Expression (13) also provides a sufficient condition for CSSD defined purely in terms of the transform function, u. This result is recorded below:

Lemma 5. *If u is (strictly) supermultiplicative over $(0,1)$, then F (strictly) CSSD dominates G, i.e.,*

$$\text{for all } s,t \in (0,1],\ u(st)(>) \geq u(s)\,u(t). \text{ then } F \text{ (strictly) CSSD dominates } G. \tag{14}$$

Proof. If the condition of the Lemma is satisfied, then clearly, expression (13) is satisfied, implying that the condition for CSSD given in Theorem 1 is satisfied. □

Both Lemma 4 and Lemma 5 beg the questions of (a) how to interpret their conditions and (b) whether their conditions are satisfied by standard distributions. In the next section, we address (a) and show that the conditions in Lemma 4 and Lemma 5 both hinge on the "geometric convexity" of the transform function. Geometric convexity formalizes and generalizes the intuition developed in Section 2. Section 11 answers (b) by showing that geometric convexity orders most common "textbook" size-indexed statistical distribution families.

4.3 Geometric convexity

The basic results from the theory of geometric convexity required to develop the stochastic order relations are provided below.

Definition 5. Let I and J be subintervals of $(0,\infty)$, then a continuous function $\phi : I \to J$ is *geometrically convex* if, for all $s,t,\in I$ and $\alpha \in (0,1)$,

$$\phi(s^{\alpha} t^{1-\alpha}) \leq \phi(s)^{\alpha}\,\phi(t)^{1-\alpha}. \tag{15}$$

If the weak inequalities are replaced with strict inequalities, we will say that ϕ is *strictly geometrically convex*. Geometric concavity is defined analogously by reversing the inequality in (15).

The following lemma, which is essentially specialized and simplified statement of Theorem 1 in Finol and Wójtowicz (2000) and Lemma 2.1.1 in Niculescu and Persson (2004), provides the basic characterizations required.

Lemma 6. *Let I and J be subintervals of $(0,\infty)$ let $\phi : I \to J$ be a continuous increasing function. Then*

(i) the following statement are equivalent.
   - *(a) ϕ is* geometrically convex.
   - *(b) The conjugate function to ϕ, $\hat{\phi} : \log(I) \to \Re$ defined by $\hat{\phi}(y) = \log\circ\phi\circ\exp(y)$ is continuous, convex, and increasing.*

(ii) if $I = (0,1]$ and $\phi(1) = 1$, conditions (a) *and* (b) *are equivalent to ϕ being supermultiplicative, i.e.,*

$$\forall (s,t) \in (0,1]\times(0,1], \quad \phi(st) \geq \phi(s)\,\phi(t). \tag{16}$$

These statements remain valid if we replace "convex" with "concave," "super" with "sub" and reverse the inequalities.

Proof. Theorem 1 in Finol and Wójtowicz (2000) establishes all of the results except the assertions that $\hat{u}$ is increasing. This result simply follows from the fact that the we have added the hypothesis that ϕ is increasing and thus conjugate map, which involves the composition of ϕ with strictly increasing functions is also increasing. □

If ϕ is differentiable, then the following lemma provides a simple test for verifying geometric convexity.

Lemma 7. *Suppose ϕ is differentiable, then ϕ is (strictly) geometrically convex if and only if*

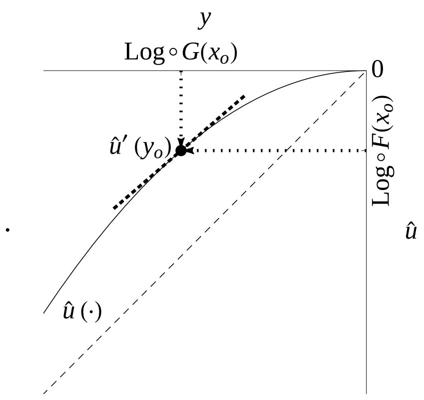


Figure 2: *Parametric representation of the conjugate transform function $\hat{u}$.* Each point in the graph of the conjugate transform function $(y, \hat{u})$ represents an ordered pair $(\log \circ G(x), \log \circ F(x))$. The graph is produced by varying x over $(\underline{x}, \bar{x}]$

the elasticity function R defined by

$$R(t) = \frac{\phi'(t)\,t}{\phi(t)},$$

is (strictly) increasing.

Proof. The result follows from observations offered after Theorem 1 in Finol and Wójtowicz (2000). □

Thus, like convexity, geometric convexity has a fundamental geometric characterization: the conjugate of ϕ, $\hat{\phi}$, defined in Lemma 6, is convex, and a derived differential characterization, provided by Lemma 7: $t \hookrightarrow t\phi'(t)/\phi(t)$ is increasing.

These two characterizations are the ones we will apply to the transform function, u, to verify geometric convexity. The function R determining geometric convexity in Lemma 7 simply represents the elasticity of the transform function, u, with respect to t, the quantile at which u is evaluated. To see this, note that we can write R as

$$R(t) = \frac{\frac{d}{dt}\log(u(t))}{\frac{d}{dt}\log t}. \tag{17}$$

The conjugate transform function, $\hat{u}$, at first glance, appears more difficult to interpret. However, as asserted in the introduction, $\hat{u}$ can be thought of simply as u plotted on a log–log scaled grid. To see this, perform the substitution $t = G(x)$ in expression (17), this yields,

$$R \circ G(x) = \frac{(\log \circ u(G(x)))'}{(\log \circ G(x))'} = \frac{(\log \circ F(x))'}{(\log \circ G(x))'}.$$

Thus, $R \circ G(x)$ represents the slope of the curve parameterized by x and defined by

$$y = \log \circ G(x), \quad \hat{u} = \log \circ F(x), \quad \underline{x} < x \leq \bar{x}.$$

This parametric log–log representation is depicted in Figure 2. Using these characterizations, the interpretation of the geometric convexity of the transform function is apparent: R being increasing implies increasing elasticity in t, i.e., that the proportionate response in the quantile of F produced by a change in a quantile of G, $G^{-1}(t)$, must be increasing in t. Intuitively, geometric convexity requires that, if F is stochastically better than G, its superiority over G must become more pronounced at higher quantiles; if F is stochastically worse than G, its inferiority must be less pronounced at higher quantiles of G. Expression (17) also shows that geometric convexity neither implies nor is implied by convexity. Geometric convexity depends on the ratio between a strictly decreasing function, $(\log(t))' = 1/t$, and $(\log(u))' = u'/u$. The transform function u being convex does not imply that u'/u is increasing unless u is also logconvex. Similarly, even if u is concave and thus $(\log(u))' = u'/u$ is decreasing, because the denominator in (17) is also decreasing, quantile elasticity, R, can be increasing.

4.4 Geometric convexity and CSSD

Using Lemma 6 the following result is immediate.

Theorem 2. *Suppose that F and G are an admissible pair of distribution functions; Let $u = F \circ G^{-1}$, then if u is (strictly) geometrically convex, F (strictly) CSSD dominates G.*

Proof. Condition (14) of Lemma 5 is sufficient for CSSD. By Lemma 6, u satisfies condition (14) if and only if u is geometrically convex. □

A slightly less obvious application of geometric convexity to the weaker sufficient condition for CSSD provided in Lemma 4 is stated and proved below.

Theorem 3. *Suppose that F and G are an admissible pair of distribution functions; Let $u = F \circ G^{-1}$, then if u is (strictly)* geometrically convex on average*, i.e., the average function U defined by equation* (10)*, is (strictly) geometrically convex, F (strictly) CSSD dominates G.*

Proof. By Lemma 7, U being (strictly) geometrically convex implies that $t \hookrightarrow (U'(t)t)/U(t)$ is (strictly) increasing. Equation (11), implies that

$$\frac{U'(t)t}{U(t)} = \frac{u(t)}{U(t)} - 1. \tag{18}$$

The right-hand side of this equation is the sufficient condition for CSSD given in Lemma 4. □

Under the elasticity interpretation provided in Section 4.3, geometric convexity on average requires that the ratio between the quantile elasticity at quantile t and the average quantile elasticity over $[0,t]$ to be uniformly increasing. Similarly, as can be seen by the formulation of CSSD provided by expression (12), CSSD dominance simply requires that the ratio between average quantile elasticity over $[0,t]$ and elasticity at t is maximized at the highest quantile, $t = 1$.

5 Properties of the dominance relations

Suppose that F and G are an admissible pair of distributions and that $u = F \circ G^{-1}$, then if u is (strictly) geometrically convex, we will say that *F (strictly) geometrically dominates G*. If U is (strictly) geometrically convex, we will say that *F (strictly) geometrically on average dominates G*. When F dominates G in the (average) geometric convexity ordering, we will write ($F \underset{ga}{\succcurlyeq} G$) $F \underset{g}{\succcurlyeq} G$. Thus the geometric dominance, and geometric dominance on average, like the CSSD defined earlier, define relations between random variable and their associated distribution functions.

The first and most transparent property of CSSD, geometric dominance, and geometric dominance on average relations is that they are invariant to common increasing transformations of the compared variables.

Lemma 8. *Suppose that $\tilde{X} \overset{d}{\sim} F$ and $\tilde{Y} \overset{d}{\sim} G$ form an admissible pair with support $[\underline{x}, \bar{x}]$, and associated transform function $u = F \circ G^{-1}$. Let $\Phi : [\underline{x}, \bar{x}] \to \Re$ be a strictly increasing function. Then the random variables, $\Phi(\tilde{X}) \overset{d}{\sim} F_\Phi$, $\Phi(\tilde{Y}) \overset{d}{\sim} G_\Phi$ form an admissible pair and*

$$\begin{aligned} &\tilde{X} \text{ CSSD (geometrically) (geometrically on average) dominates } \tilde{X} \text{ if and only if} \\ &\qquad\qquad \Phi(\tilde{X}) \text{ CSSD (geometrically) (geometrically on average) dominates } \Phi(\tilde{X}), \end{aligned}$$

with the same equivalence holding for the strict versions of the relations.

Proof. $\Phi(\tilde{X})$ and $\Phi(\tilde{X})$ clearly form an admissible pair. $F_\Phi = F \circ \Phi^{-1}$ and $G_\Phi = G \circ \Phi^{-1}$. Therefore, the quantile transform function associated with $\Phi(\tilde{X})$, u_Φ is given by

$$\begin{aligned} u_\Phi = F_\Phi \circ G_\Phi^{-1} &= (F \circ \Phi^{-1}) \circ (G \circ \Phi^{-1})^{-1} \\ &= (F \circ \Phi^{-1}) \circ (\Phi \circ G^{-1}) = F \circ (\Phi^{-1} \circ \Phi) \circ G^{-1} = F \circ G^{-1} = u. \end{aligned}$$

Because all three order relations depend only on the quantile transform function u, and the transform function is unaffected by Φ, order relations between $\Phi(\tilde{X})$ and $\Phi(\tilde{Y})$ are equivalent to order relations between $\tilde{X}$ and $\tilde{Y}$. □

Lemma 8 has two noteworthy consequences. First, any property of the of the distribution functions (or random variables) being compared that is not invariant to increasing transformations of both of the compared variables—e.g., compact supports, finite expectations—cannot be a necessary or sufficient condition for any of the dominance relations. Second, if a dominance relation is established between two distributions, the same relation must also hold between any two distributions produced by scaling or shifting these two distributions by a common factor. Later, this result will be useful for charactering dominance relations between textbook distributions.

We now turn to the less obvious question of the order properties of these dominance relations. As the next lemma, Lemma 9, reports, geometric dominance and geometric dominance on average are preorders over admissible distribution functions. However, CSSD is not transitive and thus not a preorder. The order properties of geometric dominance and geometric dominance on average simply follow from the fact that, for these orders, order comparison is based on the convexity of functional compositions in log-scaled space. Since these compositions preserve convexity, they preserve order. The best way to understand the failure of transitivity of CSSD is through inspection of expression (13). From this expression, we see that CSSD dominance is consistent with violations of the supermultiplicative condition for some values of s and t provided that violations are "integrated out." However, violations that are small enough to integrate out in one comparison of distributions may not be small enough to average out in another.[10]

Lemma 9. *(i) The geometric dominance and average geometric dominance relations are preorders over the set of admissible distributions. (ii) Geometric dominance implies geometric dominance on average and both geometric dominance and geometric dominance on average imply CSSD dominance. (iii) The CSSD dominance relation is not a preorder over the set of distributions because CSSD dominance is not transitive.*

Proof. See Supplement S1.

The failure of transitivity for CSSD would be a very significant defect if CSSD were viewed as a relation used to define rational choice over risky prospects. However, the application intended for CSSD is very different. The "selectors" do not use CSSD to rank alternatives; they use the realized values of the random prospects. Rather, CSSD is a tool for signing the comparative statics resulting from their selections. Thus, CSSD's functional application in economic arguments is closer to the typical application the MLR order than it is to stochastic dominance. In this role, the main consequence of intransitivity is to prevent the use of transitivity arguments to verify CSSD. In practice, this is probably less of a drawback than the rather complex and opaque conditions CSSD imposes on the structure of the compared distributions. Despite the failure of CSSD to define a preorder, for the sake of brevity, when we collectively refer to the three relations—geometric dominance, geometric dominance on average, and CSSD—we will simply call them "selection-conditioned orders."

We next show that, although geometric dominance and geometric dominance on average are preorders, neither is a partial order. It is possible for F and G to be distinct distributions yet $F \underset{g}{\succcurlyeq} G$ and $G \underset{g}{\succcurlyeq} F$ ($F \underset{ga}{\succcurlyeq} G$ and $G \underset{ga}{\succcurlyeq} F$) When this occurs, we will say that F and G are *geometrically equivalent (on average)*. If F and G, F CSSD dominates G and G also CSSD dominates F, we will say that the pair is *CSSD equivalent*.

If an admissible pair of distributions is equivalent, stochastic dominance is reflected in their probabilities of selection rather than their expected values conditioned on selection. In general, neither geometric dominance nor geometric dominance on average are necessary conditions for CSSD dominance. However, as the next lemma demonstrates, geometric and average geometric dominance are both necessary and sufficient conditions for CSSD equivalence. Moreover, CSSD equivalence imposes very strong conditions on distribution functions: they must be related by a power transform.

[10]See Supplement S2 for an example of intransitivity.

Lemma 10. *For an admissible pair of distribution functions, F and G, the following statements are equivalent: (i) F and G are geometrically equivalent. (ii) F and G are geometrically equivalent on average. (iii) F and G are CSSD equivalent. (iv) $F(x) = G(x)^p$ for some $p > 0$.*

Proof. See Supplement S1.

When p in part (iv) of the Lemma is a positive integer, the Lemma is very intuitive. In this case, F is the distribution of the maximal order statistic resulting from p independent draws from G. Thus, one can think of G as competing against p independent copies of itself for selection. Each of the p copies of G is selected if its value exceeds the values of the one draw from G and the other $p-1$ copies, i.e., when its value exceeds the value of p independent draws from the G distribution. This is exactly the same condition that a draw from G must satisfy for selection. Thus, the expected value conditioned on selection for each of the p copies is the same as G's expected value. Since the expected value of draws from F conditioned on selection equals the average of the selection-conditioned values of the copies, the selection-conditioned expectation of G is the same as the selection-conditioned expectation of F. Of course, if p is large, draws from F are much more likely to be selected than draws from G, but the selection-conditioned distributions of F and G are identical.

This result has two significant implications: First, it applies that our results, which compare selection-conditioned distributions of two random variables, extend to multiple random variables under the assumptions that the random variables are independent draws from two different distributions and conditioning is with respect to the draw being the maximum draw. The second implication is that, if we compare distributions that are different powers of the same underlying distribution, these distributions are selection equivalent under all three relations. For example, all distributions of the form $F(x) = x^p$, $p > 0$ and $x \in [0,1]$ are CSSD equivalent. A less obvious example of CSSD equivalence is provided by the extreme-value type distributions with the same shape but different location or scale parameters.

6 Probability densities and the MLR order

We aim to characterize the extent to which our order relations can be applied to the standard distributions used in economics and finance research. Typically, these distributions have absolutely continuous distribution functions with respect to Lebesgue measure and thus are characterized by their probability density functions. Moreover, one of the two stochastic orders we aim develop for selection-conditioned distributions, the MLR order, is typically defined in terms of probability density functions. Thus, to characterize these density functions we turn our attention to distributions with well-defined and tractable probability density functions.

Definition 6. An admissible pair of distribution functions F and G are *regularly related* if

(i) $F(0) = G(0) = 0$.

(ii) F and G are absolutely continuous with respect to Lebesgue measure on $(\underline{x},\bar{x})$.

(iii) On $(\underline{x},\bar{x})$, F and G have density functions, f and g that are strictly positive and continuous.

Regularity implies that u is differentiable and thus Lemma 7 can be used to verify that u is geometrically convex. It also implies that F and G have probability density functions. Under regularity, Lemma 7 permits us to produce a mapping between the properties of the transform function, u, and the properties of the underlying distribution and density functions of the random variables generating u. This mapping is provided by the next result, Theorem 4. In the case of geometric dominance, verifying dominance reduces to simply examining the ratio between the *reversed hazard rates* of the compared distributions, r_F and r_G, where $r_F = f/F$ and $r_G = g/G$.

Theorem 4. *Suppose that F and G are regularly related and $u = F \circ G^{-1}$. Let r_F and r_G represent the reversed hazard rates of F and G respectively. Then (i) u is (strictly) convex if and only if $x \hookrightarrow f(x)/g(x)$ is (strictly) increasing over $(\underline{x},\bar{x})$, i.e., if and only if F (strictly) dominates G in the MLR order. (ii) u is (strictly) geometrically convex, i.e., F (strictly) dominates G in the geometric dominance order, if and only if $x \hookrightarrow r_F(x)/r_G(x)$ is (strictly) increasing over $(\underline{x},\bar{x})$.*

Proof. See Supplement S1.

An immediate and interesting consequence of Theorem 4 is that a geometric dominance relation between distributions is equivalent to their selection-conditioned distributions being ordered by MLR.

Lemma 11. *If F and G are regularly related, F (strictly) geometrically dominates G if and only if F (strictly) dominates G under the competitive selection conditioned MLR ordering.*

Proof. Theorem 4 shows that the condition for geometric dominance, when it can be expressed in terms of probability densities, is identical to the condition for competitive selection MLR dominance derived in Lemma 2. □

Not only are the implications of MLR for selection-conditioned comparisons identical to the implications of MLR for unconditional comparisons, the nature of the structural constraints on the underlying distributions imposed by these orders are also quite analogous. Geometric dominance is in the same class of order relations as MLR—order relations defined by the quantile transform.[11] MLR can be expressed as the ratio, $f/g = F'/G'$, being increasing; geometric dominance can be expressed as the ratio $(\log \circ F)'/(\log \circ F)' = r_F/r_G$ being increasing. MLR dominance is equivalent to the transform function, u being convex; Geometric dominance is equivalent to u being geometrically convex. MLR dominance implies unconditional stochastic dominance; geometric dominance implies selection-conditioned stochastic dominance.

[11]The other order classes are the class of orders defined than by a cone of functions, for example stochastic dominance, and the class defined by the distribution transform $F^{-1} \circ G$, for example the van Zwet skewness order (Lehmann and Rojo, 1992).

7 Origins of geometric dominance

Our primary aim is to uncover the relation between conditional and unconditional stochastic orderings. However, we need to develop some understanding of how conditional dominance between distributions can arise in general, i.e., in the absence of stochastic dominance. Because, CSSD is not even an order relation, the question of the distributional characteristics which lead to CSSD dominance in general is difficult to frame. Geometric dominance on average, being related to the average properties of the distribution function, cannot be directly connected to the shape of the underlying distributions. In contrast, as we show in the next theorem, geometric dominance, even absent any stochastic dominance relation between the compared distributions, places very strong and simple restrictions on the compared distributions.

Theorem 5. *Suppose that F and G are an admissible pair of distributions and let $u = F \circ G^{-1}$. Suppose that F strictly geometrically dominates G.*

(i) If, on some neighborhood of $\underline{x}$, $F(x) < G(x)$, then for all $x \in (\underline{x}, \bar{x})$, $F(x) < G(x)$, and thus F strictly stochastically dominates G.

(ii) If, on some neighborhood of $\underline{x}$, $F(x) > G(x)$, then either

- *(a) $F(x) > G(x)$ for all $x \in (\underline{x}, \bar{x})$ and thus the geometrically dominated distribution, G, strictly stochastically dominates the geometrically dominant distribution, F, or*
- *(b) There exists a point $x^o \in (\underline{x}, \bar{x})$ such that for all $x \in (\underline{x}, x^o)$, $F(x) > G(x)$ and for all $x \in (x^o, \bar{x})$, $F(x) < G(x)$, i.e., the geometrically dominant distribution crosses the geometrically dominated distribution once from above.*

Sketch of proof. The formal proof of this result is presented in Supplement S1. The intuition behind the proof of this result presented here nicely illustrates how the "geometry" of geometric dominance restricts the behavior of distribution functions. At quantiles where the two distribution functions, F and G, cross, i.e., points where $F(x) = G(x)$, the conjugate transform function meets the diagonal line, $\hat{u}(y) = y$ depicted in Figure 2. Points below the diagonal map into values of the distribution functions at which $F(x) < G(x)$ and points above the diagonal map into points at which $F(x) > G(x)$. Strict geometric dominance implies strict convexity of the transform function. Convexity imposes strong restrictions how often the conjugate transform function can cross the diagonal. The strict convexity of the conjugate function implies that, if the conjugate transform function starts below the diagonal, as it will if F is stochastically dominant, it can only meet the diagonal once. However, the conjugate transform function must meet the diagonal at the upper endpoint of its domain, $y = 0$, which corresponds to the point where both distribution functions equal 1. Thus, the conjugate transform function cannot meet the diagonal at any other point and hence must lie below the diagonal for all $y < 0$. This case corresponds to part (i). If the conjugate function starts out above the diagonal, strict convexity implies that it can cross the diagonal at most twice. Again, because one of these crossings must occur at the upper endpoint, the conjugate transform crosses the diagonal at most once from above before reaching the endpoint. One crossing from above before reaching the endpoint corresponds to part (ii.b) and no crossing before reaching the endpoint corresponds to part (ii.a). □

In essence, Theorem 5 divides geometric dominance relations into three possible configurations: In the first case, developed in part (i) of the Theorem, the geometrically dominant distribution is stochastically dominant. In the second, developed in part (ii.a), the geometrically dominant distribution is stochastically dominated. In the third case, developed in part (ii.b), the geometrically dominant distribution is in some since dispersion increasing. Because our aim is to relate unconditional stochastic dominance to stochastic dominance conditioned on selection, our focus in the subsequent analysis will be on the first two cases.

8 Selection preservation

We first consider case (i) of Theorem 5. In this case, selection preserves the ordering of distributions: when one distribution is unconditionally better than another, it is also better conditioned on selection. When selection preservation takes the strong form of geometric dominance, preservation imposes sharp, standard, and fairly easy to verify restrictions on the properties of the unconditional distributions.

Theorem 6. *Suppose that F and G are an admissible pair of distributions and let $u \circ G = F$. Suppose that F strictly stochastically dominates G. If F strictly geometrically dominates G, then (i) The transform function, u, is strictly convex, i.e., (ii) F strictly MLR dominates G.*

Proof. See Supplement S1.

Theorem 6 shows that, when geometric dominance preserves the stochastic dominance ordering, it implies MLR ordering. This result is easy to establish under the assumptions that (a) the quantile transform function is twice continuously differentiable over $(0,1)$, (b) is twice differentiable at $t = 0$, and (c) $u'(0) > 0$. These assumptions are not innocuous as they are more restrictive than the conditions for regularity we have imposed. In fact, (b) never satisfied in the case of selection reversal considered next. However, they permit a simpler and more intuitive derivation than the derivation developed under more the general conditions in Supplement S1. The intuitive derivation works as follows: Consider the elasticity function R defined in Lemma 7. Its derivative, which exists by (a), can be expressed as

$$R'(t) = \frac{1}{u(t)}\left(R(t)\left(\frac{u(t)}{t} - u'(t)\right) + t\,u''(t)\right), \quad t \in (0,1). \tag{19}$$

Applying LH́ôpital's rule, which is justified by (b) and (c), shows that $\lim_{t\to 0} R(t) = 1$. Because strict geometric convexity implies that R is strictly increasing, $R(t) > 1$, $t \in (0,1)$, which, by the definition of R, implies that $u(t)/t - u'(t) < 0$. The strict geometric convexity condition provided by Lemma 7 can only be satisfied if $R' > 0$. Inspection of equation (19) shows that $R' > 0$ only if $u'' > 0$, i.e., u is strictly convex. Thus, by Theorem 4.ii, the stochastically dominant distribution is MLR dominant. In essence, the drag on quantile elasticity produced by the average value of the quantile transform function being less than its marginal value has to be countered by increases in the marginal value, requiring convexity of the transform function and thus MLR dominance of the stochastically dominant distribution.

9 Selection reversal

Example 1 showed that a point mass at 0 can induce selection reversal. In fact, as the next lemma shows, a point mass at $\underline{x}$ always strictly strengthens selection dominance regardless of the distributions being compared or which conditional dominance ordering is employed.

Lemma 12. *Suppose let F_o be and admissible distribution function with $F_o(\underline{x}) = 0$ such that F_o (CSSD) (geometrically on average) (geometrically) dominates G; let δ_0 represent the distribution function for a point mass at 0. let $F = p\,\delta_0 + (1-p)\,F_o$, $p \in (0,1)$, then F strictly (CSSD) (geometrically on average) (geometrically) dominates G.*

Proof. See Supplement S1.

The point mass case analyzed in Lemma 12 is, as we will show in Section 12.2, not just a statistical illustration but rather the equilibrium outcome of an effort-bidding model (Baye et al., 1993, 1996), and more generally of asymmetric bidder all-pay auction models.

Given the discussion in Section 2, Lemma 12 might not be too surprising. The next result is perhaps more surprising: it shows that "something like" a point mass at $\underline{x}$ is the *only* way to induce reversal: a point mass per se is not required but, for reversal to occur, the ratio between the left tail weights assigned by the two distributions must approach infinity as the left-tail shrinks to the lower bound of the distributions' support.

Theorem 7. *Suppose that F and G are an admissible pair of distributions and let $u \circ G = F$. Suppose that F is strictly stochastically dominated by G. Then,*

(i) if F CSSD dominates G, $\limsup_{x \to \underline{x}} \frac{F(x)}{G(x)} = \infty$.

(ii) If in addition, F strictly geometrically dominates G, then (a) for all $x \in (\underline{x}, \bar{x})$, the function $x \hookrightarrow \frac{F(x)}{G(x)}$ is strictly decreasing and (b) $\lim_{x \to \underline{x}} \frac{F(x)}{G(x)} = \infty$.

Proof. See Supplement S1.

Consistent with the intuition provided in Section 4.3, selection reversal requires the stochastically dominated but geometrically dominant distribution's inferiority to be most pronounced at low quantiles of the stochastically dominant distribution. Although always "worse" than the stochastically dominant distribution, at higher quantiles, its inferiority is less pronounced; thus, the elasticity of the transform function is increasing as required by geometric dominance.

Note also that the conditions for selection preservation and reversal are not symmetric, selection reversal does not imply that the distributions are reverse MLR ordered. The intuition for this result is simply that the explosion of the left-tail ratio around the lower endpoint of the distributions implies that at the lowest quantile $t = 0$, the derivative of the quantile transform function, u', is either infinite or undefined. Thus, the intuitive argument provided below Lemma 6 fails because its assumptions are not satisfied. This asymmetry is also not surprising given Theorem 5, which shows geometric convexity imposes asymmetric restrictions on distribution functions when the geometrically dominant distribution is stochastically dominant as opposed to stochastically dominated.

10 Preservation of geometric dominance under operations

In this section, we develop a few of the basic properties of the geometric dominance order. Some of the results are required for the derivations (outlined in Section 11) of dominance relations between specific parametric distributions. Others are included to provide further insight into the nature of geometric dominance and its relation to the MLR order, its unconditional analog.

10.1 Upscalings and upshifts

Suppose that G is an absolutely continuous distribution function with support $(-\infty,\infty)$ and that the density of g is continuous over $(-\infty,\infty)$. Then if $F(x) = G(x-c)$, $c > 0$, we will say that F is a *c-upshift* of G and call F the *upshifted distribution* and G the *original distribution*, in which case, $\tilde{X} \overset{d}{\sim} G$, implies that $\tilde{X} + c \overset{d}{\sim} F$, Similarly, suppose that G is an absolutely continuous distribution function with support $[0,\infty)$ and that the density of g is continuous over $(0,\infty)$. If $F(x) = G(x/s)$, $s > 1$ we will say that F is an *s-upscaling* of G and call F the *upscaled distribution* and G the *original distribution*, in which case, $\tilde{X} \overset{d}{\sim} G$, implies that $s\tilde{X} \overset{d}{\sim} F$, As the next lemmas show, for upshifts and upscalings, geometric dominance only depends on the generalized convexity properties of the reversed hazard rate of the original distribution.

Lemma 13. *All c-upshifts of G (strictly) geometrically dominate G if and only if the reversed hazard rate function of G, $r = g/G$ is (strictly) logconcave.*

Proof. See Supplement S1.

Lemma 14. *All s-upscalings of G (strictly) geometrically dominate G if and only if the reversed hazard rate function of G, $r = g/G$ is (strictly) geometrically concave.*

Proof. See Supplement S1.

How "strict" are the conditions for the preservation of order under upshifts and upscaling relative to each other and relative to the analogous conditions for the preservation of MLR ordering? The condition for upshifts and upscalings to preserve MLR dominance is logconcavity of the probability density function (Theorem 1.C.18 and Theorem 1.C.22: Shaked and Shanthikumar, 1994)? The upshift and upscaling conditions in Lemmas 13 and 14, as well as the logconcavity of the probability density, all imply the logconcavity of the *distribution* function, i.e., that the distribution has a decreasing reversed hazard rate.[12] For decreasing functions, logconcavity implies geometric concavity. Thus, the condition on the reversed hazard rate for upshifts in Lemma 13 is more stringent than the condition imposed for upscalings in Lemma 14. In fact, logconcavity of the reversed hazard rate implies logconcavity of the probability density.[13] Thus, the condition imposed for upshifts in Lemma 13, is more stringent than the analogous condition for upshifts imposed by MLR. In contrast, the geometric concavity condition on the reversed hazard rate imposed for upscalings in Lemma 14 neither implies or is implied by logconcavity of the probability density.[14] Thus, the condition for upscalings preserving geometric dominance and the condition for upscalings to preserve MLR order do not nest.

10.2 Convolutions and mixtures

Like MLR dominance, geometric dominance is not generally preserved by mixtures, i.e., if F_i geometrically dominates G_i for all $i = 1, 2, \ldots n$, it need not be the case that convex combinations of F_i geometrically dominate convex combinations of G_i. Geometric dominance has weaker preservation properties under convolutions than MLR. While MLR is not generally preserved by convolutions, it is preserved by convolutions of the compared distributions with the same distribution if that distribution has a logconcave density.[15] This is not the case for geometric dominance. The failure of convolution with a logconcave density to generally preserve geometric dominance follows because a logconcave density can be very close to being flat over some portion of its support. Convolving a distribution with such a logconcave density flattens out the likelihood ratio of the compared distributions. Flattening the likelihood ratio can cause the elasticity condition for geometric dominance to fail.[16]

[12]For a proof that both the upscaling and upshift conditions imply a decreasing reversed hazard rate see Lemmas S1-7 and Lemma S1-8 in Supplement S1; the implication for logconcave density functions is well known (e.g., Theorem 1: Baricz, 2010).

[13]See Lemma S1-9 in Supplement S1.

[14]See Examples-S1.1 and S1.2 in Supplement S1.

[15]For a more complete discussion of the preservation of MLR ordering under convolutions and mixtures, see Shaked and Shanthikumar (Chapter 1.C, 1994)

[16]Examples supporting these assertions are available upon request.

11 Distributions

Using the results developed above, particularly Lemmas 13 and 14, we provide, in Table 1, an accounting of the effects of selection on standard textbook distributions. Typically, these distribution families can be represented by two parameters: a "shape" parameter and a "size" parameter. In most cases, under standard parameterizations, for a fixed shape parameter, the map between the size parameter and the and the distribution is order preserving in the stochastic dominance order. In a few cases, standard textbook parameterizations are order reversing. In these cases, we reparameterize the distribution family to make it order preserving. When the size parameter of one distribution in an ordered family is larger than the size parameter of another, we call the distribution with the larger size parameter the *upsized distribution* and we call the other distribution the *original distribution*. Location/scale families of distributions, are examples of size-ordered families in which each pair of distributions in the family are related by an upscaling or an upshift.

For location–scale families, By Lemma 8, we need only consider the effect of selection on the "standardized" distribution of the family, e.g., for the Normal distribution family, the Normal(0,1) distribution. Thus, for location–scale families, we can simply apply Lemmas 13 and 14 to the standardized distribution. For families which are not location–scale, we apply the more foundational results developed earlier in the paper.

We ask the question of whether, holding the shape parameter fixed, increasing the size parameter implies that the upsized distribution selection dominates the original distribution under our three selection-conditioned dominance relations—CSSD, geometric dominance on average, and geometric dominance. If, under at least one of these relations, the upsized distribution is always strictly dominant, we say that dominance is preserved and report the strongest ordering condition satisfied. If increasing the size parameter leads to the upsized distribution being strictly dominated under at least one of the selection-conditioned orders, we say that dominance is reversed and report the strongest ordering condition under which the reversal occurs. If upsized distribution is selection equivalent to the original distribution we say that selection is *neutralized*. Of course, it is possible that, conditioned on selection, the upsized distribution might not be related to the original distribution under any the selection-conditioned order relations. However, for the textbook distributions we examine, this possibility is never realized. The results of this investigation are reported in Table 1. The sometimes tedious, sometimes trivial, calculations underlying these results are deferred to Supplement S3.

Table 1: **Selection preservation and reversal**

Distribution	**Parameters** *Size*	*Type*	*Shape (fixed)*	**Effect of selection on dominance** *Dominance of the upsized distribution is*
Normal $F[x] = \frac{1}{2} - \frac{1}{2}\,\mathrm{Erf}\left(\frac{\mu - x}{\sqrt{2}\sigma}\right)$ $x \in (-\infty, \infty)$	μ	location	$\sigma > 0$	*Preserved* Strictly geo. dominant
Logistic $F[x] = \frac{1}{1+e^{-\frac{x-\mu}{s}}}$ $x \in (-\infty, \infty)$	μ	location	$s > 0$	*Preserved* Strictly geo. dominant
Laplace $F[x] = {}^{1}\!/_{2} \exp\left(\frac{x-\mu}{s}\right)$ if $x \geq 0$ $F[x] = 1 - {}^{1}\!/_{2} \exp\left(-\left(\frac{x-\mu}{s}\right)\right)$ if $x < 0$ $x \in (-\infty, \infty)$	μ	location	$s > 0$	*Preserved* Avg. geo. dominant, strictly CSSD dominant but not geo. dominant
Gumbel $F[x] = \exp\left(-e^{-\frac{x-\mu}{s}}\right)$ $x \in (-\infty, \infty)$	μ	location	$s > 0$	*Neutralized* Selection equivalent
Gamma $F[x] = (1/\Gamma(\alpha)) \int_0^{x/s} z^{\alpha-1} e^{-z}\, dz$ $x \in [0, \infty)$	s	scale	$\alpha > 0$	*Preserved* Strictly geo. dominant
Generalized Exponential[a] $F[x] = \left(1 - e^{-\frac{x}{s}}\right)^{b}$ $x \in [0, \infty)$	s	scale	$b > 0$	*Preserved* Strictly geo. dominant
Weibull[a] $F[x] = 1 - e^{-\left(\frac{x}{a}\right)^{\lambda}}$ $x \in [0, \infty)$	$a > 0$	scale	$\lambda > 0$	*Preserved* Strictly geo. dominant
Pareto[b] $F[x] = 1 - \left(\frac{x_m}{x}\right)^{\frac{\mu}{\mu-1}}$ $x \in [x_m, \infty)$	$\mu > 1$	neither	$x_m > 0$	*Preserved* Strictly geo. dominant

Table 1: (continued)

Distribution	**Parameters** *Size*	*Type*	*Shape (fixed)*	**Effect of selection on dominance** *Dominance of the upsized distribution is*
Kumaraswamy $F[x] = 1 - (1 - x^{\alpha})^b$ $x \in [0, 1]$	$\alpha > 0$	neither	$b > 0$	*Preserved* if $b > 1$ Strictly geo. dominant *Neutralized* if $b = 1$ selection equivalent *Reversed* for $b < 1$ Strictly geo. dominated
Lognormal $F[x] = \frac{1}{2} - \frac{1}{2}\,\mathrm{Erf}\left(\frac{\mu - \log(x)}{\sqrt{2}\sigma}\right)$ $x \in [0, \infty)$	μ	scale	$\sigma > 0$	*Preserved* Strictly geo. dominant
Fréchet $F[x] = \exp\left(-\left(\frac{x}{s}\right)^{-\alpha}\right)$ $x \in [0, \infty)$	s	scale	$\alpha > 0$	*Neutralized* Selection equivalent
Log-logistic $F[x] = \frac{1}{1+\left(\frac{x}{\alpha}\right)^{-\beta}}$ $x \in [0, \infty)$	$\alpha > 0$	scale	$\beta > 0$	*Preserved* Strictly geo. dominant
Gompertz[c] $F[x] = 1 - \exp\left(\eta\,\left(1 - e^{x/s}\right)\right)$ $x \in [0, \infty)$	$s > 0$	scale	$\eta > 0$	*Preserved* Strictly geo. dominant if $\eta \geq 1$ Strictly CSSD dominant but not geo. dominant and not always geo. dominant on average if $\eta < 1$

[a] The Exponential distribution is a special case of the Weibull distribution and Generalized Exponential, the case where $\lambda = 1$ for the Weibull and $b = 1$ for the Generalized Exponential.
[b] Standard parameterization changed to ensure that increasing the size parameter produces stochastically larger distributions. Replacing μ with $\alpha/(1-\alpha)$ produces the standard parameterization.
[c] Standard parameterization of distribution modified to ensure that increases in the scale parameter up scales rather than down scales. Replacing s with $1/b$ produces the standard parameterization.

The most striking observation about Table 1 is that size parameter shifts lead to strict geometric dominance for the upsized distribution in the vast majority of cases. Because, geometric convexity implies selection-conditioned MLR dominance, and as shown in Theorem 6, MLR dominance is a necessary condition for geometric dominance when the geometrically dominant distribution is stochastically dominant, these result show that, for the overwhelming majority of distributions surveyed, the MLR ordering is robust to selection.

Given the results thus far, most of the exceptional cases where selection is not preserved or partially preserved are quite easy to rationalize. The Gumbel and Fréchet distributions are extreme value type distributions and, as explained in Section 5, such distributions are always equivalent conditioned on selection. For two distributions on our list, the Laplace and Gompertz, upsizing does not result in geometric dominance but some other form of dominance is preserved. For the Gompertz distribution when $\eta < 1$, the upscaled distribution is CSSD dominant but not geometrically dominant and sometimes not even geometrically dominant on average. The failure of geometric dominance in this case is also easy to explain. When the Gompertz shape parameter, η is less than 1, the upscaled distribution is not MLR dominant, thus, by Theorem 6 it cannot be geometrically dominant. In contrast, for the Laplace distribution, the upshifted distribution is MLR dominant. As MLR dominance of the upshifted distribution is a necessary but not sufficient condition for geometric dominance, the failure of geometric dominance is, of course, consistent with our earlier results. In the case of the Laplace distribution, the failure of geometric convexity results from nonconvexity of the conjugate transform function around the quantile of the original distribution that maps into the mean of the upshifted distribution. Averaging smooths out the nonconvexity and thus the upshifted distribution is always geometrically dominant on average.

The only distribution in the table in which selection completely reverses dominance is the Kumaraswamy distribution for shape parameters $b < 1$. In this case, the upshifted distribution is strictly MLR dominant but strictly geometrically dominated. What accounts for selection causing a complete reversal of dominance in this case? Perhaps the best way to understand this reversal case is to compare it with the "normal" case where selection is preserved. What better exemplar can one find of the normal case than the Normal distribution itself. Thus, below, in Figure 3, we plot the transform, u and conjugate transform, $\hat{u}$ functions of the Kumaraswamy and Normal distributions. In both cases $u = F \circ G^{-1}$ maps the relation between a distribution, G, and its upsizing, F. The graph of the transform function can be thought of as a parametric plot where each point on the graph is given by $(G(x), F(x))$, for some point x in the support of the distributions. Thus, the abscissa in Figure 3 is labeled G and the ordinate is labeled F.

From Panels A and B of the figure, we see that u, in both the Kumaraswamy and Normal cases, is convex and lies below the diagonal, showing that in both cases the upscaled distribution is stochastically and MLR dominant. However, as the graph makes apparent, the nature of MLR dominance in the two cases is very different. In the case of the Kumaraswamy, the convexity of u is concentrated in the very lowest quantiles. Thus, the convexity of u is not increasing a rate proportional to proportional quantile increases. As discussed in Section 4.3, this pattern indicates geometric nonconvexity. In contrast, for the Normal distribution, the convexity of u is increasing. Thus, reversal is expected in the Kumaraswamy case but not in the

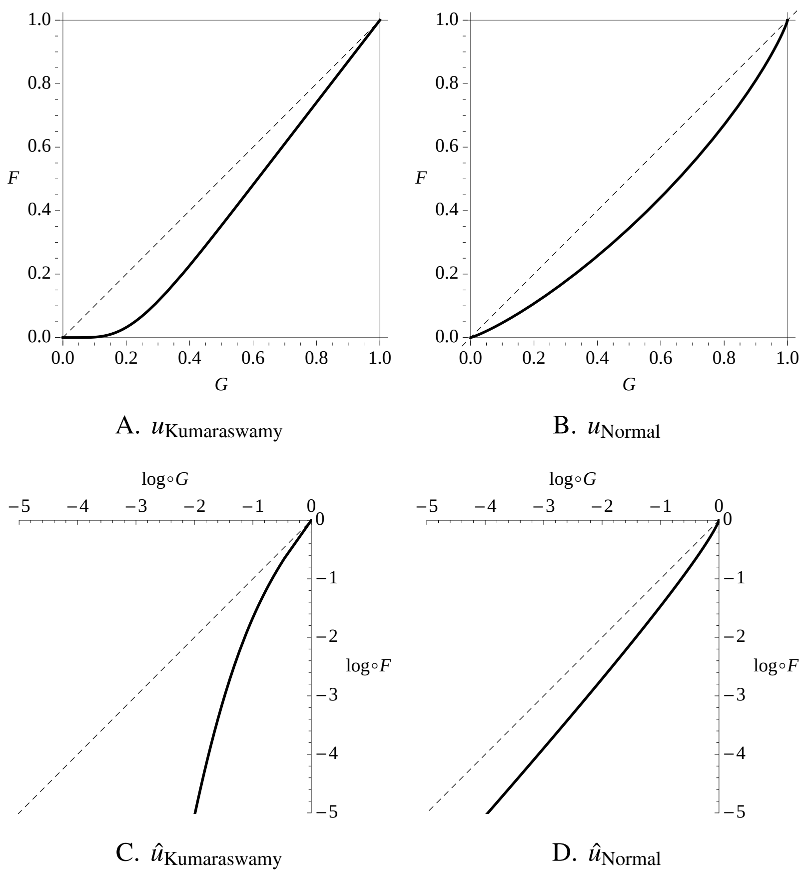


Figure 3: *Plots of the transform, u, and conjugate transform, û, functions for the Kumaraswamy and Normal Distributions, in both cases F is an upsizing of G.* In the Kumaraswamy case, the common shape parameter is $b = 1/8$ and the size parameter for F is $\alpha = 8$ and for G is $\alpha = 1$. In the Normal case, the common shape parameter is $\sigma = 1$ and the size parameter for F is $\mu = 0.40$ and for G is $\mu = 0$.

Normal case. This expectation is verified by Panels C and D which plot the conjugate transform functions. The conjugate transform in the Normal case is convex and thus the upsized Normal is geometrically dominant while the conjugate transform in the Kumaraswamy case is *concave* indicating that, in fact, the stochastically smaller original distribution geometrically dominates the upsized distribution.

12 Applications

In this section, we illustrate, through concrete examples, how the order relations developed above can be applied to selection conditioning in microeconomic models. The applications show that these selection orders can be used to draw important policy-relevant implications from standard microeconomic models, provide conditions for robust qualitative inference for simple statistical decision problems, and generate selection-robust model predictions.

12.1 Admissions tournaments

Consider a university that receives applications from two groups of applicants, group A and group B. A concrete example of such groups might be state-school graduates and independent-school graduates. Admission is based on applicant quality, $\tilde{Q} \geq 0$, which is itself determined by applicant investment in effort, $e \geq 0$, to acquire knowledge, as well as a random shock with an absolutely continuous distribution which captures the uncertain relation between effort to acquire knowledge and knowledge acquisition. The cost of effort for applicants from group i, $i = A, B$ is given by e_i/a_i, $a_i > 0$. Thus, a is a measure of ability, with more able applicants having lower effort costs. Except for incorporating heterogeneity, and changing the parametric assumptions on the structure of random shocks, the framework is identical to the classic Lazear and Rosen (1981) tournament model.

At date 0, applicants make their effort choice. Admission is decided at date 1, and the admission decision is determined by head-to-head competition between pairs of applicants from the two groups. The winner of the competition is the applicant with the highest quality unless both applicant's quality equals 0, in which case neither applicant is admitted. At date 2, admitted students receive a score from the university based on their academic performance, $\tilde{S}$. Their score depends both on their quality and a university-specific shock, $\tilde{F}$, independent of the quality of the student. For simplicity, we assume that the relationship between the score, applicant quality, and the university-specific shock, is given by $\tilde{S} = \tilde{F}\,\tilde{Q}$. Our task is to derive a model under which the score of admitted students identifies which group of applicants are more able.

As will be apparent in the sequel, the absolute continuity of the shock terms ensures that ties occur with probability 0 and thus they will be ignored. The value of being admitted to applicants is normalized to 1. We assume that the structural relation between applicant quality, $\tilde{Q}$, effort, and the random shock is given as follows: The quality of applicants from group i, $\tilde{Q}_i$ is given by

$$\tilde{Q}_i = e_i \tilde{Z}_i, \quad \tilde{Z}_i \overset{d}{\sim} \text{Exponential}(\beta),\ i = A, B, \tag{20}$$

where $\tilde{Z}_A$ and $\tilde{Z}_B$ are independent random variables. The multiplicative error structure is a departure form the literature which typically assumes that an additive random shock generated by a distribution which is symmetric, absolutely continuous, centered at 0, and supported by $(-\infty, \infty)$. We depart from the literature because the typical assumption leads, as Lazear and Rosen note, to the non-existence of pure strategy equilibria for some parameterization of the model. When agents are heterogeneous, the non-existence problems are even more severe and ensuring pure-strategy equilibria requires imposing complex restrictions on effort costs function of the two groups (Gürtler and Kräkel, 2010). As will be seen shortly, our exponential multiplicative shock specification avoids these problems.[17]

By the symmetry of the game, it is sufficient to consider the best-responses of group A applicants and then exploit symmetry to produce the best responses of group B applicants. So consider a group A applicant, given that applicants in group B choose effort level e_B, the probability that the applicant will win the admissions tournament is the probability that $\tilde{Q}_A > \tilde{Q}_B$. Using equation (20) we see that the probability of admission is given by $\mathbb{P}[\tilde{Z}_B/\tilde{Z}_A < e_A/e_B]$. Next note that $\tilde{Z}_B/\tilde{Z}_A \overset{d}{\sim} x/(1+x)$, $x \geq 0$. Thus, using the normalization of the value of admission to 1, the assumed effort costs, and the discussion above, we see that the utility of a group A agent, v_A, is given by

$$v_A(e_A, e_B) = \begin{cases} \frac{e_A}{e_A+e_B} - \frac{e_A}{a_A} & \text{if } e_A + e_B > 0, \\ 0 & \text{if } e_A + e_B = 0. \end{cases}$$

Similarly, for a group B agent,

$$v_B(e_A, e_B) = \begin{cases} \frac{e_B}{e_A+e_B} - \frac{e_B}{a_B} & \text{if } e_A + e_B > 0, \\ 0 & \text{if } e_A + e_B = 0. \end{cases}$$

No positive level of effort is a best reply to the other applicant choosing zero effort and choosing zero effort is also never a best reply to the other applicant choosing 0 effort. Thus, in any Nash equilibrium, both applicants choose positive effort levels. Because, each applicant's utility is strictly concave in effort, optimal positive effort choices are determined by the first-order conditions:

$$\frac{\partial}{\partial e_A} v_A = \frac{e_B}{(e_A + e_B)^2} - \frac{1}{a_A} = 0, \tag{21}$$

$$\frac{\partial}{\partial e_A} v_B = \frac{e_A}{(e_A + e_B)^2} - \frac{1}{a_B} = 0. \tag{22}$$

It is easy to verify that these equations have unique solution for positive effort levels given by

$$e_i^* = a_i \left(\frac{a_A\, a_B}{(a_A + a_B)^2} \right) = \tfrac{1}{4}\, a_i \, \frac{\mathscr{H}(a_A, a_B)}{\mathscr{A}(a_A, a_B)}, \tag{23}$$

where $\mathscr{A}$ represents the arithmetic mean and $\mathscr{H}$ the harmonic mean of ability. Thus, for all choices of the ability parameters, a pure strategy equilibrium exists in which high-ability agents exert more effort. Note also that the ratio between the harmonic arithmetic means is inversely related to the differences in ability of the two groups. Thus, this simple model captures the two basic results of heterogeneous-agent tournament theory: more able agents choose higher effort levels and differences in ability attenuate the effort incentives of both high and low-ability agents.

[17] Actually any distribution with a concave CDF will yield a unique pure-strategy equilibrium. Nor does the scale parameter have to be the same for both types. Different scale parameters simply translate into different marginal productivities of effort which is a perfect substitutes for marginal effort costs. However, alternative parameterization would not yield the simple closed-form solution developed below.

If an applicant from group i is admitted, the quality, $\tilde{Q}_i$, of the applicant will equal $e_i^* \tilde{Z}_i$. An admitted students academic performance or "score", $\tilde{S}_i$, will thus equal $e_i^* \tilde{Z}_i \tilde{F}$. An applicant from group i is admitted if and only if her quality exceeds the quality of the rival, i.e., $e_i^* \tilde{Z}_i > e_j^* \tilde{Z}_j$, $j \neq i$. Thus, the selection-conditioned score of group i applicants is given by

$$\mathbb{E}[\tilde{S}_i | \, \text{Admitted}] = \mathbb{E}[e_i^* \tilde{Z}_i \tilde{F} \,|\, e_i^* \tilde{Z}_i > e_j^* \tilde{Z}_j]. \tag{24}$$

Using these observations, we can verify that the model makes a selection-robust prediction about the quality of applicants from the two groups. To see this note that,

$$\begin{aligned} & a_i > a_j \overset{(1)}{\iff} e_i > e_j \overset{(2)}{\iff} \\ & \mathbb{E}[e_i^* \tilde{Z}_i | e_i^* \tilde{Z}_i > e_j^* \tilde{Z}_j] > \mathbb{E}[e_j^* \tilde{Z}_j | e_j^* \tilde{Z}_j > e_i^* \tilde{Z}_i] \overset{(3)}{\Longrightarrow} \\ & \mathbb{E}[e_i^* \tilde{Z}_i \tilde{F} \,|\, e_i^* \tilde{Z}_i > e_j^* \tilde{Z}_j] > \mathbb{E}[e_j^* \tilde{Z}_j \tilde{F} \,|\, e_j^* \tilde{Z}_j > e_i^* \tilde{Z}_i]. \end{aligned} \tag{25}$$

The first equivalence follows from the model solution given in equation (23). The second follows because $e_i^* > e_j^*$ if and only if $e_i^* \tilde{Z}_i$ is an upscaling of $e_j^* \tilde{Z}_j$. As shown by Table 1 upscalings of exponential distributions result in geometric dominance. By Theorem 2 geometric dominance implies selection dominance. The third implication follows because CSSD is preserved by common shocks (Lemma 1). Using the same chain of logic, we see that

$$a_i \leq a_j \Longrightarrow \mathbb{E}[e_i^* \tilde{Z}_i \tilde{F} \,|\, e_i^* \tilde{Z}_i > e_j^* \tilde{Z}_j] \leq \mathbb{E}[e_j^* \tilde{Z}_j \tilde{F} \,|\, e_j^* \tilde{Z}_j > e_i^* \tilde{Z}_i]. \tag{26}$$

Thus expressions (25), (26), and (24) imply that

$$a_i > a_j \iff \mathbb{E}[\tilde{S}_i | \, \text{Admitted}] > \mathbb{E}[\tilde{S}_j | \, \text{Admitted}].$$

Thus, identifying the selection-conditioned mean sore of selected students permits qualitative inference about the ability of unselected applicants from the two pools: applicants from the group whose admitted members perform better at university have greater ability. Note also that the quality of the more able group strictly stochastically dominates the quality of the less able group. Thus, the probability of being admitted for the more able group is always greater than for the less able group. Hence, a greater probability of admission for one group combined with better university performance of the other is not consistent with non-discriminatory quality-based admissions policies.

12.2 Effort bidding admissions contests

In Section 12.1, competition for admission was modeled in a tournament framework in which effort is determined first and the effect of effort on applicant quality is mediated by an exogenous shock. In this section, assuming the same structural relations between effort, ability, and applicant quality, we model admission selection in Baye et al. (1993, 1996) effort bidding framework. In this framework, agents choose effort distributions which directly determine applicant quality without the mediation of an endogenous shock. Applicants select probability distributions over non-negative effort levels. The applicant with the highest realized effort wins admission. Thus, in the notation developed in Section 12.1, $\tilde{Q}_i = \tilde{e}_i$, $i = A,B$, $\tilde{e}_i \overset{d}{\sim} F_i$ As in Section 12.1 the highest quality applicant is selected for admission.[18] Thus, in the Baye et al. (1993, 1996) framework, admission is determined by an all-pay auction in which applicants bid effort. Using the ability/effort relation postulated in Section 12.1, the payoffs to agents of each type, w_i, are thus given by

$$w_A(e_A|F_B) = F_B(e_A) - \frac{e_A}{a_A},$$
$$w_B(e_B|F_B) = F_A(e_B) - \frac{e_B}{a_B}.$$

An equilibrium is a pair of effort distributions (F_A, F_B) satisfying the condition that each effort level in the support of the effort distributions is a best reply to the effort distribution selected by applicants of the other type. Assume, without loss of generality that $a_A \geq a_B$, i.e. type A applicants are more able. The following distribution pair (F_A^*, F_B^*) is an equilibrium:

$$F_A^*(e) = \min[e/a_B, 1],$$
$$F_B(e) = p\,\delta_0(e) + (1-p)\,F_A^*(e), \text{ where } p = (a_A - a_B)/a_A,$$

where δ_0 is the distribution function for a point mass at 0. To see that (F_A^*, F_B^*) is an equilibrium, note that

$$v_A(e_A|F_B^*) = \frac{a_A - a_B}{a_A} - \frac{a_B}{a_A}\max\left[0, \frac{e_A}{a_B} - 1\right],$$
$$v_B(e_B|F_A^*) = -\max\left[\frac{e}{a_B} - 1, 0\right].$$

Thus, over the common support of F_i, $i = A,B$, which equals $[0, a_B]$ the payoffs to both applicants of type A and applicants of type B are constant and their payoffs are strictly decreasing off the support for feasible effort levels (i.e., when $e > a_B$). Expected effort is given by

$$\mathbb{E}[\tilde{e}_i^*] = \tfrac{1}{2}\,a_i\,\frac{a_B}{a_A}, i = A,B.$$

Thus, as in tournament model, expected effort is increasing in ability and decreasing in the difference between the ability of the applicants.

Because we have already illustrated, in Section 12.1, how to incorporate common random shocks, we assume in this section that the performance of students after admission simply equals student quality, which under the effort-bidding framework is determined by effort. Thus, the academic performance of admitted students, S in the notation of Section 12.1, equals their effort. Hence, the expected performance of admitted applicants of type $i = A,B$ is given by

$$\mathbb{E}[\tilde{S}_i|\,\text{Admitted}] = \mathbb{E}\left[\tilde{e}_i^*|\,\tilde{e}_i^* > \tilde{e}_j^*\right], i \neq j.$$

It is clear that the unconditional equilibrium effort distribution for type A, F_A^*, strictly stochas-

[18] In Baye et al. (1993, 1996) and all-pay auction models, equilibrium "bid" (effort in our context) distributions are continuous and thus tie-bids occur with zero probability under a tie-breaking rule that offers admission to each candidate with equal probability in the event of a tie. Hence, to shorten our exposition, we ignore tie bids.

tically dominates type B's equilibrium effort distribution F_B^*. However, because $F_B^* = p\,\delta_0 + (1-p)\,F_A^*$, $p \in (0,1)$, Lemma 12, implies that F_A^* is strictly geometrically dominated by F_B^*. This implies by Theorem 2 and definition of CSSD (Definition 2) that the expected academic performance of type B admitted students is better than the academic performance of type A admitted students.

Because F_A^*, strictly stochastically dominates F_B^* the probability that an applicant of type A will be selected is also strictly higher than the probability that an applicant of type B will be selected. Thus, the Oxford admissions paradox, discussed in the introduction, can be rationalized under the assumption of non-discriminatory selection by the Baye et al. (1993, 1996) effort bidding framework but not by the classical Lazear and Rosen (1981) tournament framework: under non-discriminatory admissions policies, students from the more able group, A, will alway have a higher probability of admission and lower average post-admissions performance than students from the less able group, B.

12.3 Identifying the sign of average treatment effects under self selection

In this section, we apply the results developed above to the problem of identifying the sign of treatment effects in the presence of self-selection bias. This is a classic inference problem examined in a number of papers(e.g., Manski, 1990). Consider two treatments, treatment 1 and treatment 2. The effect of treatment $i = 1,2$ on a given patient is given by

$$\tilde{Y}_i = m_i + \tilde{U}_i, \quad i = 1,2.$$

where $\tilde{U}_1$ and $\tilde{U}_2$ are independent, integrable, zero-mean, identically distributed random variables with a common distribution function. This distribution function, F_o, has a continuous probability density function and is supported by the real line. The distributions of $\tilde{Y}_1$ and $\tilde{Y}_2$ can be expressed as $F_1(y) = F_o(y - m_1)$ and $F_2(y) = F_o(y - m_2)$. Thus, $F_1(y) = F_2(y - (m_1 - m_2))$ and $F_2(y) = F_1(y - (m_1 - m_2))$. Hence, in the language of Section 10, if $m_1 > m_2$ then F_1 is an upshift of F_2 and if $m_1 < m_2$ then F_2 is an upshift of F_1.

Patients select the treatment that is optimal for them and thus choose treatment i whenever $\tilde{Y}_i > \tilde{Y}_j$, $i \neq j$. The expected effect of treatment i on the patients who undertake the treatment is thus given by $\mathbb{E}[\tilde{Y}_i | \tilde{Y}_i > \tilde{Y}_j]$. Identifying the sign of the average treatment effect involves determining which treatment, if applied to the entire population, would produce the best results, i.e., determining whether the sign of the difference in unconditional mean effects of the treatments can be identified from the sign of the difference in mean effects conditioned on selection. Thus, our problem reduces to determining the conditions under which

$$\operatorname{sgn}[m_1 - m_2] = \operatorname{sgn}\left[\mathbb{E}[\tilde{Y}_1 | \tilde{Y}_1 > \tilde{Y}_2] - \mathbb{E}[\tilde{Y}_2 | \tilde{Y}_2 > \tilde{Y}_1]\right]. \tag{27}$$

The very simple self-selection model developed above is can be mapped into to the Roy model (Roy, 1951). Using the notation and structure in the exposition of the Roy model provided by Heckman (2008), the two "treatments" are two market sectors in which workers can be employed. Workers have sector-specific skills that can only be applied in one of the two markets. Workers are heterogeneous and their sector-specific skills in each of the two sectors are represented by the random variables $\tilde{T}_i$, $i = 1, 2$. The wage rate in sector i is represented by π_i. Thus, if a worker works in sector i, the worker's wage, $\tilde{W}_i$, is given by $\tilde{W}_i = \pi_i \tilde{T}_i$. Hence, wage-maximizing workers will choose to work in sector i if $\tilde{W}_i > \tilde{W}_j$. Since worker wages in a sector are only measured if a worker chooses to work in a sector, measured average wages in sector i equal selection-conditioned wages. The Roy model assumes that

$$\log(\tilde{W}_i) = \log(\pi_i) + \mu_i + \tilde{U}_i, \quad i, = 1, 2.$$

where μ_i is a constant representing the log-mean of $\tilde{T}_i$, and $\tilde{U}_1$, $\tilde{U}_2$ are zero-mean jointly normally distributed random variables. Letting $\tilde{Y}_i = \log(\tilde{W}_i)$ and $m_i = \log(\pi_i) + \mu_i$, we see that apart from the assumed stochastic structure, the treatment effect problem we pose is isomorphic to the problem considered by the Roy model. In the Roy context, being able to sign the treatment effect from the conditional treatment effect amounts to being able to infer form the fact that measured wages are higher in one sector, that average wages, $\pi \mathbb{E}[\tilde{T}_i]$ are also higher.

The difference between the stochastic structure we consider and the structure considered in the Roy model is that the Roy model assumes that the error terms are Normally but non-identically distributed. We do not restrict the error terms to being Normal but we do assume that the error terms in the two sectors are identically distributed.[19]Also, as will be apparent from the analysis below, we aim to find conditions under which a much stronger identification hypothesis can be verified: observing any monotone function of measured wages can identify the sign of the difference in average wages even in the presence of sample truncation.

To develop these conditions, first consider the trivial case where $m_1 = m_2$. In this case $\tilde{Y}_1$ and $\tilde{Y}_2$ are identically distributed and thus $\mathbb{E}[v(\tilde{Y}_2)|\tilde{Y}_2 > \tilde{Y}_1] = \mathbb{E}[v(\tilde{Y}_1)|\tilde{Y}_1 > \tilde{Y}_2]$, where v is any (measurable and integrable) valuation function. When $m_1 > m_2$, then F_1 is an upshift of F_2 and so, by the definition of CSSD, if upshifts engender strict CSSD dominance, equation (27) will be satisfied. Similarly, if $m_1 > m_2$, then F_2 is an upshift of F_1 and so, by the definition of CSSD (Definition 2), if upshifts engender strict CSSD dominance, equation (27) will be satisfied. In fact, because CSSD implies dominance not only of the expectation of the conditional treatment effects of the CSSD distribution but also of any increasing non-constant valuation function, v, of the treatment effect, if upshifts engender strict CSSD dominance, then

$$\operatorname{sgn}[m_1 - m_2] = \operatorname{sgn}\left[\mathbb{E}[v(\tilde{Y}_1)|\tilde{Y}_1 > \tilde{Y}_2] - \mathbb{E}[v(\tilde{Y}_2)|\tilde{Y}_2 > \tilde{Y}_1]\right]. \tag{28}$$

Theorem 2 shows that strict geometric dominance implies strict CSSD dominance, Lemma 13

[19] By using Lemma 1 and assuming that the error term for each treatment is the sum an identically distributed independent shock and a common shock, these results could be extended to account for correlated errors. However, even in this extended framework, it would sill be the case that the marginal distributions of the error terms are identical.

shows that all upshifts engender strict geometric dominance if and only if the reversed hazard rate of the common error distribution, F_o, is logconcave. Thus, if the error distribution has a logconcave reversed hazard rate, Equation (28) and *a fortiori* equation (27) holds. As verified in Appendix S3, the Normal and Logistic distributions have logconcave reversed hazard rates. Many other potential error distributions, e.g., the hyperbolic secant distribution, also satisfy this logconcavity property. In addition, as presented in Table 1 and demonstrated in Appendix S3, the Laplace distribution, while not satisfying the conditions for strict geometric dominance, satisfies the condition that upshifts always engender strict CSSD dominance. Thus, in all of these cases, (28) is satisfied which *a fortiori* implies that (27) is satisfied. The satisfaction of equation (28) implies that the difference in average treatment effects can be identified from the sign of the difference between any increasing non-constant function of the selection-conditioned treatment effects. When upshifts engender geometric dominance even more can be said: Lemma 2 shows that geometric dominance implies selection-conditioned MLRP dominance and, thus, in this case, the sign of the difference in average treatment effects can be identified even from truncated samples (Lemma 3).

The results developed thus far are sufficient conditions for identification of the sign of the difference between average treatment effects under the assumption that the error terms under the two treatments are independent and identically distributed. Some of these results are admittedly not surprising. E.g., the fact that for normally distributed error distributions, (27) is satisfied is well known and implicit in the analysis of a number of papers. Is the extension of this result to an entire class of distributions (distributions with logconcave reversed hazard rates), a large class of valuation functions (increasing and non-constant), and to truncated observations surprising? The answer depends on whether one views the preservation of the sign of treatment effects as obvious or surprising. However, even if one views preservation as obvious, the results in this paper are also useful because they provide a “recipe” for constructing “surprising” counterexamples to the “obvious” equality postulated in equation (27) using quite “innocuous” distribution functions.

Lemma 13 shows that logconcavity of the reversed hazard rate is the key to ensuring that upshifts engender geometric dominance. Hence, constructing a distribution that falsifies (27) amounts to finding an error law with a reverse hazard rate that exhibits significant logconvexity. This is not hard to do even if one restricts attention to standard distributions.

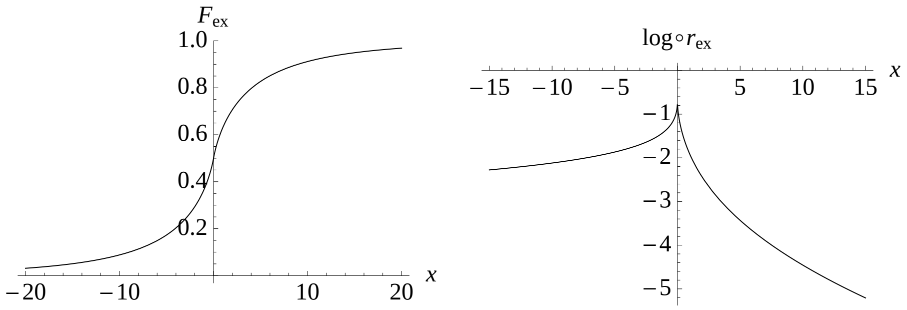
Consider this *example distribution*, a Generalized Normal distribution with shape parameter $1/2$.[20] The distribution, F_{ex} is defined by

$$F_{\text{ex}}(x) = \frac{1}{2} + \operatorname{sgn}(x)\left(\frac{1}{2} - \frac{1}{2}e^{-\sqrt{|x|}}\left(1+\sqrt{|x|}\right)\right), \quad f_{\text{ex}}(x) = \frac{1}{4}\exp\left(-\sqrt{|x|}\right).$$

The example distribution, F_{ex}, is symmetric, zero-mean, unimodal, and has moments of all

[20]The generalized normal distribution is also called the “exponential power distribution” or “generalized error distribution.” See Saralees (2005) for an extensive analysis of the Generalized Normal distribution.

orders (Saralees, 2005).[21] This distribution and the log of its reversed hazard rate are plotted below in Figure 4. Although the reversed hazard rate for F_{ex} is not logconvex, restricted to either



A. PDF of F_{ex} B. Log of the reversed hazard rate of F_{ex}

Figure 4: *Generalized Normal distributions with shape parameter* 1/2

the positive nor negative half-line it is logconvex. Thus, it appears to be a good candidate for a counterexample to equality (27). In fact, it is easy to verify numerically that when $m_1 = 2$ and $m_2 = 0$, and thus the difference between the unconditional treatment effects is 2, the difference between the conditional treatment effects is given by

$$\mathbb{E}[\tilde{Y}_1 | \tilde{Y}_1 > \tilde{Y}_2] - \mathbb{E}[\tilde{Y}_2 | \tilde{Y}_2 > \tilde{Y}_1] \approx -0.2318,$$

and equality (27) fails. Hence, even when the errors under the two treatments are identically distributed and independent, the error distribution is symmetric and has moments of all orders, and there are no confounding unobservable (or observable) variables, selection can reverse the direction of inference. In the Roy model context, this example shows, under the same conditions, that an increase in the wage rate in one sector can reduce expected measured wages in that sector.

12.4 Selection in first-price auction models

Our aim in this section is to determine the conditions under which the winning bids of the stronger bidders MLR dominate the winning bids of the weaker bidders in asymmetric auction models. These models are used to model auctions in which there are known ex ante asymmetries between bidders. Such situations are quite common, e.g., takeovers where one bidder has a toehold stake, procurement auctions with domestic (favored) and foreign (unfavored) bidders, contract auctions were some bidders are part of a bidding cartel (e.g., Cantillon, 2008; Pesendorfer, 2000). When the auctioneer is not a public agency, e.g., sealed-bid divestiture sales of corporate divisions (Baker and Wruck, 1989), bids are typically private information, and only the terms of the paid (winning) bid are publicly observed. Our analysis will rely on Maskin and Riley (2000), which determines the unconditional bid distribution of strong and

[21] F_{ex} is, like the log-normal distribution, heavy tailed but not fat tailed.

weak bidders in asymmetric first-price, private-value auctions. As discussed in the conclusion, second-price and English auctions are also amenable to an analysis by a stochastic ordering based on "deselection dominance." However, the development of deselection dominance is beyond the scope of this paper.

Maskin and Riley provide assumptions under which the equilibrium distributions of bids for both the strong and weak bidders will be supported by the same interval, $[b_*, b^*]$ and are twice continuously differentiable. Over the interior this interval, the ratio of the reversed hazard rates for the bid distribution of the strong, r_s and weak, r_w, bidders verify the following condition:

$$\frac{r_s(b)}{r_w(b)} = \frac{\phi_s(b) - b}{\phi_w(b) - b}, \quad b \in (b_*, b^*), \tag{29}$$

where ϕ_i is the equilibrium inverse bid function for a bidder of type i. This characterization follows from equation 3.23 in Maskin and Riley (2000), reproduced here for the reader's convenience:

$$\frac{p_s'}{p_s} = \frac{1}{H_w(p_w) - b}, \quad \frac{p_w'}{p_w} = \frac{1}{H_s(p_s) - b}. \qquad \text{(Maskin and Riley, 2000, 3.23)}$$

To see that equation (29) is a simple consequence of equation 3.23 in Maskin and Riley (2000), first note that p_i'/p_i, $i = s, w$, in their notation, is simply the reversed hazard rate of the bid distribution for type i, r_i. Next, note that H_i, in their notation, is the inverse of the value distribution for a bidder of type i. The distribution of bids for a bidder of type i is given by $p_i = F_i(\phi_i(b))$, where F_i is the value distribution for type i; thus, $H_i(p_i) = \phi_i$.

The functions $b \hookrightarrow \phi_i(b) - b$, $i = s, w$, represent bid shading—the difference between the value of the auctioned good when a bidder i bids b, given by $\phi_i(b)$, and the bid that bidder i submits. Thus, we can think of the ratio on the right-hand side of equation (29) as the *bid-shading ratio*. An increasing bid-shading ratio is equivalent to the elasticity of bid shading (with respect to the bid) being higher for the strong bidder than for the weak bidder. Equation (29) implies that the ratio between the reversed hazard rates, r_s/r_w, is (strictly) increasing if and only if the bid-shading ratio is (strictly) increasing. Theorem 4 shows that a (strictly) increasing ratio between the reversed hazard rates is equivalent to (strict) geometric dominance. With probability 1, a bidder in the first-price auction equilibrium wins the auction and thus pays her bid if and only if her bid is higher than the other bidder's bid. Thus, the distribution of winning bids is the selection-conditioned distribution of bids. Theorem 11 shows that (strict) geometric dominance of the strong bidder's bid implies that the selection-conditioned distribution of the strong bidder's bid MLR dominates the selection-conditioned distribution of the weak bidder's bid.

Moreover, Maskin and Riley also show (Proposition 3.3.ii, Maskin and Riley, 2000) that the unconditional distribution of the strong bidder's bid always strictly stochastically dominates the unconditional distribution of the weak bidder's bid. Thus, geometric dominance combined with stochastic dominance, imply, by Theorem 6.ii, that, in addition, the unconditional distribution of the strong bidder's bid MLR dominates the weak bidder's bid. These observations establish the following result.

Result 1. *Under the assumptions of Maskin and Riley (2000), the bid distribution and selection-conditioned bid distribution of the strong bidder (strictly) MLR dominates both the unconditional and selection-conditioned distribution of the weak bidder if and only if the equilibrium bid-shading ratio, $(\phi_s(b)-b)/(\phi_w(b)-b)$, is (strictly) increasing in the bid, b.*

Thus, Result 1 clearly implies that the bid-shading ratio being increasing is a sufficient condition for the expected payment by the strong bidder to exceed the expected payment by the weak bidder, i.e., that strong bidders pay more in auctions.

The test for geometric dominance provided by Result 1 can easily be applied to specific inverse bid functions in cases where closed form representations of these functions are available. For example, Kaplan and Zamir (2012) explicitly derive in detail the inverse bid functions when the strong bidder's value distribution is uniform between 0 and v_s and the weak bidder's valuation function is uniform between 0 and v_w, $v_w < v_s$. In this case, they show that the equilibrium inverse bid functions are given by

$$\left.\begin{aligned} \phi_w(b) &= \frac{2b\,(v_s v_w)^2}{(v_s v_w)^2 + b^2\,(v_s^2 - v_w^2)}, \\ \phi_s(b) &= \frac{2b\,(v_w v_w)^2}{(v_w v_w)^2 - b^2\,(v_s^2 - v_w^2)}. \end{aligned}\right\}, \quad b \in [0, (v_s v_w)/(v_s + v_w)]$$

These inverse bid functions imply that the bid-shading ratio equals

$$\frac{\phi_s(b)-b}{\phi_w(b)-b} = \left(\frac{v_s^2 v_w^2 + b^2\left(v_s^2 - v_w^2\right)}{v_s^2 v_w^2 - b^2\,(v_s^2 - v_w^2)}\right)^2, \quad b \in [0, (v_s v_w)/(v_s + v_w)],$$

which is clearly strictly increasing in b, verifying the increasing bid-shading ratio condition and thus, by Result 1, verifying the strict MLR dominance of both the conditional and unconditional bid distribution of the strong bidder.[22]

13 Conclusion

This paper considers the effect of selection on inference when selection is based on ranking. This question is of considerable relevance to any economic of statistical problems characterized by rank-based selection and agent heterogeneity. Heterogeneity might be generated by the effect of a treatment, as in the treatment-effects framework, or be the product of differing endowed agent characteristics, as in heterogeneous-agent auction and tournament models. Rank-based selection might be selection by the agents themselves as in self selection or rank-based

[22] In fact, Kaplan and Zamir (2012) also consider an auction with a binding reserve price under the same distributional assumptions. It can be shown that the strict geometric dominance of the strong bidder's bid also holds in this setting.

selection by a third party based on performance, as in tournaments and auctions. Given the vast output of research on such problems, the paper clearly deals with an important topic. The question answered by the paper with respect to such problems is the scope of valid inference of dominance relations from unconditional heterogeneity to selection-conditioned heterogeneity and vice versa. The paper answers this question by using stochastic order relations to provide necessary and sufficient conditions for such directional inference. Given that much of economic research is centered on determining the directional effects of changes in model parameters on agent behavior, directional characterizations are valuable. Given that the distributional relations developed in this paper are, in some cases, quite analogous to the MLR order, a relation ubiquitously postulated in theoretical microeconomics, and satisfied by the distributions typically used in parametric microeconomic models of agent behavior, these relations are clearly not too restrictive or complex for application. In fact, the paper demonstrates applicability by applying its results to specific models and parametric distributions, with the number of models and distributions considered limited by reader exhaustion rather than exhaustion of the applications.

Moreover, this paper's research program is extensible. The most obvious and natural extension is to consider conditioning on other selection events based on the realizations of the compared random variables. For example, one might consider, "deselection-conditioned dominance," where deselection denotes the event that the realized value of a random variable is less than the realized value of the compared random variable. This extension has some obvious applications. For example, in two bidder English and second price auctions, the minimum reservation bid determines the price at which the auctioned good is sold. Thus, an analysis of deselection dominance could determine the distributional conditions under which a bidder, having a stochastically dominant valuation, would pay more or less in expectation for the auctioned good. Another interesting extension would be to consider noisy selection. Selection contingent, for example, on the random variable being selected by quantile response utility function.

References

Athey, Susan, "Monotone comparative statics under uncertainty," *Quarterly Journal of Economics*, 2002, *117* (1), 187–223.

Baker, George P. and Karen H. Wruck, "Organizational changes and value creation in leveraged buyouts," *Journal of Financial Economics*, 1989, *25* (2), 163 – 190.

Baricz, Árpád, "Geometrically concave univariate distributions," *Journal of Mathematical Analysis and Applications*, 2010, *363* (1), 182–196.

Baye, Michael R., Dan Kovenock, and Casper G. de Vries, "Rigging the lobbying process: an application of the all-pay auction," *The American Economic Review*, 1993, *83* (1), pp. 289–294.

_ , _ , and _ , "The all-pay auction with complete information," *Economic Theory*, 1996, *8* (2), 291–305.

Cantillon, Estelle, "The effect of bidders' asymmetries on expected revenue in auctions," *Games and Economic Behavior*, 2008, *62* (1), 1 – 25.

Finol, C. E. and M. Wójtowicz, "Multiplicative properties of real functions with applications to classical functions," *Aequationes Mathematicae*, 2000, *59* (1-2), 134–149.

Garner, Richard, "State school students "get better degrees" than private school pupils with same A-levels," *Independent*, September 2015.

Gazette, Oxford, "Statistical Information on the University of Oxford," 2016.

Gürtler, Oliver and Matthias Kräkel, "Optimal tournament contracts for heterogeneous workers," *Journal of Economic Behavior & Organization*, 2010, *75* (2), 180–191.

Han, Song, "Discrimination in Lending: Theory and Evidence," *The Journal of Real Estate Finance and Economics*, 2004, *29* (1), 5–46.

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

Heckman, James J., "Sample selection bias as a specification error," *Econometrica*, 1979, *47* (1), pp. 153–161.

_ , "Selection bias and self-selection," in Steven N. Durlauf and Lawrence E. Blume, eds., *The New Palgrave Dictionary of Economics*, Palgrave Macmillan, 2008.

_ **and Bo E. Honoré**, "The empirical content of the Roy model," *Econometrica*, 1990, *58* (5), pp. 1121–1149.

Kaplan, Todd R and Shmuel Zamir, "Asymmetric first-price auctions with uniform distributions: Analytic solutions to the general case," *Economic Theory*, 2012, *50* (2), 269–302.

Kirkegaard, René, "A mechanism design approach to ranking asymmetric auctions," *Econometrica*, 2012, *80* (5), 2349–2364.

Laffont, Jean-Jacques, Herve Ossard, and Quang Vuong, "Econometrics of first-price auctions," *Econometrica*, 1995, *63* (4), 953–980.

Lavoie, Marc, Gilles Grenier, and Serge Coulombe, "Discrimination and Performance Differentials in the National Hockey League," *Canadian Public Policy / Analyse de Politiques*, 1987, *13* (4), 407–422.

Lazear, Edward P. and Sherwin Rosen, "Rank-order tournaments as optimum labor contracts," *Journal of Political Economy*, 1981, *89* (5), 841–864.

Lehmann, E. L. and J. Rojo, "Invariant directional orderings," *Annals of Statistics*, 12 1992, *20* (4), 2100–2110.

Machin, Stephen, Richard Murphy, and Zeenat Soobedar, "Differences in labour market gains from higher education participation," Report to National Equality Panel 2009.

Manski, Charles F., "Nonparametric bounds on treatment effects," *The American Economic Review*, 1990, *80* (2), pp. 319–323.

Maskin, Eric and John Riley, "Asymmetric auctions," *The Review of Economic Studies*, 2000, *67* (3), 413–438.

McAfee, R Preston and John McMillan, "Auctions and bidding," *Journal of Economic Literature*, 1987, *25* (2), 699–738.

Milgrom, Paul R., "Good news and bad news: Representation theorems and applications," *The Bell Journal of Economics*, 1981, *12* (2), pp. 380–391.

Niculescu, Constantin P. and Lars-Erik Persson, *Convex Functions and Their Applications: A Contemporary Approach*, Springer-Verlag, 2004.

Pesendorfer, Martin, "A study of collusion in first-price auctions," *The Review of Economic Studies*, 2000, *67* (3), 381–411.

Quah, John K-H and Bruno Strulovici, "Comparative statics, informativeness, and the interval dominance order," *Econometrica*, 2009, *77* (6), 1949–1992.

Roy, Andrew Donald, "Some thoughts on the distribution of earnings," *Oxford economic papers*, 1951, *3* (2), 135–146.

Saralees, Nadarajah, "A generalized normal distribution," *Journal of Applied Statistics*, 2005, *32* (7), 685–694.

Shaked, Moshe and J. George Shanthikumar, *Stochastic Orders and Their Applications*, Academic Press, 1994.

Stevens, MJ, D Bhattacharya, and S Kanaya, "Are university admissions academically fair?," Working paper, University of Oxford. Forthcoming in the *Review of Economics and Statistics* 2016.

For online publication only

Supplement to

Stochastic orders and the anatomy of selection

Thomas Noe

S1 Proofs and supplementary results

S1.1 Definitions and supplementary results

In order to shorten some of the derivations in the subsequent sections, we define a few functions that will be used in many of the subsequent proofs.

Definition S1-1. Define the function, $\Pi[u]$ where u is an admissible transform function by

$$\Pi[u](t) = U(t)/u(t) = \frac{1}{t}\int_0^t \frac{u(s)}{u(t)}\,ds, \quad t \in (0,1].$$

Lemma S1-1. *For any fixed admissible quantile transform function $u = F \circ G^{-1}$, the function $\Pi[u] : (0,1] \to \Re$*

(i) is continuous;

(ii) F CSSD dominates G if and only if

$$\forall t \in (0,1), \quad \Pi[u](t) \geq \int_0^1 u(s)\,ds \equiv \Pi[u](1).$$

Proof. Part (i) follows from continuity and positivity of u over $(0,1]$ which implies the continuity and positive of its average function, U. Part (ii) is simply a rephrasing of the condition for CSSD provided by expression (12). □

S1.2 Properties of the average transform function

We require some results on the average transform function, U to produce the proofs in this part of the supplement, as well as to characterize some of the distributions studied in Supplement S3. Because these results will be used in many of the derivations, it is best to derive them together for reference rather than bury them in proofs or repeat their derivation each time they are needed. The reader might want to refer to these results on an as needed basis when reading the derivations in Section S1.3.

Definition S1-2. For admissible transform functions, u, define

$$R_U(t) = \frac{U'(t)\,t}{U(t)},$$

where U represents the average transform function defined by expression (10).

Definition S1-3. For admissible pairs of distribution functions, F and G, define

$$\mathscr{R}(x) = \frac{\int_{\underline{x}}^{x} G(z)\,f(z)\,dz}{\int_{\underline{x}}^{x} g(z)\,F(z)\,dz}, \quad x \in (\underline{x}, \bar{x}).$$

Lemma S1-2. *For any admissible transform function, u,*

$$U(t) < u(t).$$

Proof. u, being admissible, is strictly increasing, so for $t \in (0,1]$,

$$U(t) = \frac{1}{t}\int_0^t u(s)\,dx < \frac{1}{t}\,t\,u(t) = u(t).$$

□

Lemma S1-3. *If u is an admissible function, and u is (strictly) geometrically convex, then it is (strictly) geometrically convex on average.*

Proof. Let,

$$I(t) = \int_0^t u(s)\,ds.$$

By Montel's theorem (Theorem 2.4.1: Niculescu and Persson, 2004), u being (strictly) geometrically convex implies that I is (strictly) geometrically convex. Next note that $U(t) = I(t)/t$. Using the conjugate expression definition for geometric convexity given in Lemma 6.b, we see that $\hat{U}(y) = \widehat{I/t}(y) = \hat{I}(y) - y$. Because I is (strictly) geometrically convex, $\hat{I}$ is (strictly) convex. The sum of a linear function and a (strictly) convex function is (strictly) convex. Thus, $\hat{I}(y) - y = \hat{U}(y)$ is (strictly) convex. By definition, this implies that U is (strictly) geometrically convex. □

Lemma S1-4. *The average transform function, U, has the following properties:*

(i) The composition of U with G represents the probability that draws from F will be less than draws from G conditioned on the draws from G being less than x, i.e.,

$$U \circ G(x) = \frac{1}{G(x)}\int_{\underline{x}}^{x} F(z)\,dG(z) = \mathbb{P}\left[\tilde{X} \le \tilde{Y} \,|\, \tilde{Y} \le x\right].$$

(ii) If F (strictly) stochastically dominates G, then

$$U(t)(<) \le \tfrac{1}{2}t,$$

and reverse inequality holds when F is stochastically dominated by G.

Proof. Part (i) follows from a simple change of variables. Given part (i), it follows that if If $x = \bar{x}$ then $U \circ G(\bar{x})$ is simply the probability that draws from F will be less than draws from G, i.e.,

$$U \circ G(\bar{x}) = \int_0^{G(\bar{x})} u(s)\, ds = \mathbb{P}\Big[\tilde{X} \le \tilde{Y}\Big].$$

If F dominates G by stochastic dominance, then $u(t) \le t$. Thus

$$t\,U(t) = \int_0^t u(s)\, ds \le \int_0^t s\, ds = \tfrac{1}{2} t^2.$$

Dividing both sides of this expression by t establishes part (ii) with the inequality holding strictly if $F(x) > G(x), x \in (\underline{x}, \bar{x})$, i.e., F is strictly stochastically dominant. The reverse inequality follows in like fashion. □

Lemma S1-5. *Suppose that F and G are regularly related and $u = F \circ G^{-1}$.*

(i) U is (strictly) convex on average if and only if

$$x \hookrightarrow \frac{\int_{\underline{x}}^x f(z)\, G(z)\, dz}{\int_{\underline{x}}^x G(z)\, g(z)\, dz}$$

is (strictly) increasing over $(\underline{x}, \bar{x})$.

(ii) U is (strictly) geometrically convex on average if and only if

$$x \hookrightarrow \mathscr{R}(x) \quad (R \text{ defined by Definition S1-3})$$

is (strictly) increasing over $(\underline{x}, \bar{x})$.

Proof. First consider (i). By equation (11), $U' = u(t)/t - U(t)/t$. This equation of can be rewritten as

$$U'(t) = \frac{\frac{1}{2}\int_0^t \Big(u(t) - u(s)\Big)\, ds}{\int_0^t s\, ds}. \tag{S1.1}$$

Because the distributions are regularly related, u' exists and is absolutely continuous, thus

$$u(t) - u(s) = \int_s^t u'(r)\, dr.$$

Thus, expression (S1.1) can be rewritten as

$$U'(t) = \frac{\frac{1}{2}\int_0^t \Big(\int_s^t u'(r)\, dr\Big)\, ds}{\int_0^t s\, ds}. \tag{S1.2}$$

Using Fubini's Theorem to reverse the order of integration yields

$$U'(t) = \frac{\frac{1}{2}\int_0^t u'(r)\Big(\int_0^r\, ds\Big)\, dr}{\int_0^t s\, ds} = \frac{1/2\int_0^t u'(s) s\, ds}{\int_0^t s\, ds} = \frac{1/2\int_0^t u'(s) s\, ds}{1/2\, t^2} = \frac{\int_0^t u'(s) s\, ds}{t^2}.$$

Finally, note that as shown in the proof of part (i) of Theorem 4, $u' \circ G^{-1} = f/g$. Using this result and performing a change of variables shows that

$$U' \circ G(x) = \frac{\int_{\underline{x}}^x G(z)\, f(z)\, dz}{G(x)^2} = \frac{\int_{\underline{x}}^x G(z)\, f(z)\, dz}{\int_{\underline{x}}^x G(z)\, g(z)\, dz}. \tag{S1.3}$$

Because G^{-1} is strictly increasing, this expression must be strictly increasing if U' is to be strictly increasing as required by the strict convexity of U.

Now consider part (ii). Note that Lemma S1-4 and (S1.3) imply that

$$R_U \circ G(x) = \frac{G(x)\, U' \circ G(x)}{U \circ G(x)} = \mathscr{R}(x), \tag{S1.4}$$

where R_U defined in Definition S1-2 and $\mathscr{R}$ in Definition S1-3. Because G is strictly increasing, R_U is (strictly) increasing if and only if $R \circ G$ is (strictly) increasing. R_U being (strictly) increasing necessary and sufficient for U being (strictly) geometrically convex, i.e., for F to (strictly) strictly geometrically on average dominate G. □

Lemma S1-6. *Suppose that F and G are an admissible pair of distributions and let $u \circ G = F$. Suppose that F is strictly stochastically dominated by G. Then, if F strictly geometrically dominates G on average, then there exists $x^o \in (\underline{x}, \bar{x})$ such that for all $x < x^o$, the function*

$$a.\ x \hookrightarrow \frac{\int_{\underline{x}}^{x} \frac{F(z)}{G(z)}\, G(z)\, d\, G(z)}{\int_{\underline{x}}^{x} G(z)\, d\, G(z)} \text{ is strictly decreasing and } b.\ \lim_{x \to \underline{x}} \frac{F(x)}{G(x)} = \infty.$$

Proof. First, define the function

$$\hat{\beta}(y) = \hat{U}(y) - y, \quad y \leq 0.$$

First note that, because U is strictly geometrically convex, $\hat{U}$ is thus strictly convex and hence $\hat{\beta}$ is strictly convex. $\hat{\beta}$ is the conjugate transformation of $U(t)/t$. Because geometric dominance on average implies CSSD dominance, and, thus by part (i) of Theorem 7, $U(t)/t$ is not increasing everywhere. Thus, $U(t)/t$ must be strictly decreasing on some interval. The conjugate transform is continuous, smooth, and order preserving, so it must be the case that $\hat{\beta}$ is also strictly decreasing on some interval. But $\hat{\beta}$ is convex and thus, if it is strictly decreasing at any point, say y^o, it is strictly decreasing for all $y < y^o$. Thus for $y < y^o$, β is strictly decreasing.

Using the definition of the conjugate function given in Lemma 6 we can see that

$$\hat{\beta} \circ \log(t) = \hat{U} \circ \log(t) - \log(t).$$

The definition of the conjugate function implies that $U(t) = \exp \circ \hat{U} \circ \log(t)$. Thus,

$$\frac{U(t)}{t} = \exp[\hat{\beta}(\log(t))]. \tag{S1.5}$$

Because the log and exp functions are order preserving, and β is strictly decreasing when $\log t < y^o$, equation (S1.5) implies that there exists $t^o \in (0,1)$ such that for all $t < t^o$, $U(t)/t$ is strictly decreasing. Finally note that the function defined in the conclusion of the Lemma is simply the composition of $U(t)/t$ with G. Thus we have established assertion (a) of Lemma S1-6.

To prove assertion (b) of Lemma S1-6, note that monotonicity of $U(t)/t$ combined with the result in the proof of part (i) of Theorem 7, which showed that $\limsup_{t \to 0} U(t)/t = \infty$, implies that $\lim_{t \to 0} U(t)/t = \infty$. Lemma S1-2 implies that $u(t)/t \geq U(t)/t$. Hence, $\lim_{t \to 0} U(t)/t = \infty$ implies that $\lim_{t \to 0} u(t)/t = \infty$, which implies by the same argument as used in part (i) of Theorem 7, the assertion in the lemma that $\lim_{x \to \underline{x}} F(x)/G(x) = \infty$. □

S1.3 Proofs of Selected Propositions

Proof of Lemma 1. Fix a constant $z \in \Re$. Note that $x \hookrightarrow v(x)$ is increasing if and only if $x \hookrightarrow v(\Gamma(z,x))$ is increasing. Thus, the CSSD dominance of $\tilde{X}$ over $\tilde{Y}$ implies that, for all $z \in \Re$,

$$\mathbb{E}\left[v(\Gamma(z,\tilde{X}) \middle| \tilde{X} > \tilde{Y}\right] \geq \mathbb{E}\left[v(\Gamma(z,\tilde{Y}) \middle| \tilde{Y} > \tilde{X}\right]. \tag{S1.6}$$

Independence and the absolute continuity of $\tilde{Z}$ imply that

$$\begin{aligned} \mathbb{E}\left[v(\Gamma(\tilde{Z},\tilde{X}) \middle| \tilde{X} > \tilde{Y}\right] &= \int_S \mathbb{E}\left[v(\Gamma(z,\tilde{X}) \middle| \tilde{X} > \tilde{Y}\right] h(z)\, dz, \\ \mathbb{E}\left[v(\Gamma(\tilde{Z},\tilde{Y}) \middle| \tilde{Y} > \tilde{X}\right] &= \int_S \mathbb{E}\left[v(\Gamma(z,\tilde{Y}) \middle| \tilde{Y} > \tilde{X}\right] h(z)\, dz. \end{aligned} \tag{S1.7}$$

Thus, expressions (S1.6), (S1.7), and the monotonicity of the integral viewed as a linear functional imply that

$$\mathbb{E}\left[v(\Gamma(\tilde{Z},\tilde{X}) \middle| \tilde{X} > \tilde{Y}\right] \geq \mathbb{E}\left[v(\Gamma(\tilde{Z},\tilde{Y}) \middle| \tilde{Y} > \tilde{X}\right]. \tag{S1.8}$$

Since $\{\tilde{X} > \tilde{Y}\} = \{\Gamma(\tilde{Z},\tilde{X}) > \Gamma(\tilde{Z},\tilde{Y})\}$ and $\{\tilde{Y} > \tilde{X}\} = \{\Gamma(\tilde{Z},\tilde{Y}) > \Gamma(\tilde{Z},\tilde{X})\}$, expression (S1.8) implies that

$$\mathbb{E}\left[v(\Gamma(\tilde{Z},\tilde{X}) \middle| \Gamma(\tilde{Z},\tilde{X}) > \Gamma(\tilde{Z},\tilde{Y})\right] \geq \mathbb{E}\left[v(\Gamma(\tilde{Z},\tilde{Y}) \middle| \Gamma(\tilde{Z},\tilde{Y}) > \Gamma(\tilde{Z},\tilde{X})\right].$$

The proof for strict CSSD dominance is the same, and involves simply replacing the weak inequalities with strong inequalities. □

Proof of Lemma 9. To prove (i) note that we need to show that the geometric convexity relation is reflexive and transitive. If F, G, and K are three admissible distribution functions, then the relation is reflexive (i) if F dominates F, and transitive (ii) if F dominates G and G dominates K implies that F dominates K. To show this, note that, by definition, F dominates itself if the function $u = F \circ F^{-1}$ is geometrically convex. Because $u = F \circ F^{-1}$ is the identity, its geometric convexity is immediate. Now consider transitivity. Transitivity will hold whenever the functions, $u_1 = F \circ G^{-1}$ and $u_2 = G \circ K^{-1}$ being geometrically convex implies that the function $u_3 = G \circ K^{-1}$ is geometrically convex. Geometric convexity holds if and only if $\hat{u}_3 = \hat{u}(y) = \log \circ u_3 \circ \exp$ is convex. Because $u_1 \circ u_2 = F \circ G^{-1} \circ G \circ K^{-1} = F \circ K^{-1} = u_3$,

$$\hat{u}_3 = \hat{u}(y) = \log \circ u_1 \circ u_2 \circ \exp = (\log \circ u_1 \circ \exp) \circ (\log \circ u_2 \circ \exp) = \hat{u}_1 \circ \hat{u}_2.$$

Because u_1 and u_2 are geometrically convex, $\hat{u}_1$ and $\hat{u}_2$ are convex. Because, u_1 and u_2 are increasing, $\hat{u}_1$ and $\hat{u}_2$ are increasing. The composition of increasing convex functions is convex. Thus, $\hat{u}_1 \circ \hat{u}_2$ is convex. Thus, the geometric convexity relation is a preorder. Exactly the same proof, with u replaced by U, verifies that geometric convexity on average is a preorder.

To prove (ii), first note that all the assertions in the Lemma follow from Theorems 2 and 3 except the assertion that geometric dominance implies geometric dominance on average. This assertion follows from Lemma S1-3.

Now consider (iii). The CSSD is clearly reflexive; however, it is not transitive. This can be verified by a counterexample available in Supplement S2. □

Proof of Lemma 10. First note that, by Lemma S1-3, (i) implies (ii). By Theorem 3, (ii) implies (iii). Next, we show that (i) implies (iv). To see this, let $u = F \circ G^{-1}$. Then $F \underset{g}{\succcurlyeq} G$ if and only if u is geometrically convex. $G \underset{g}{\succcurlyeq} F$ if and only if u^{-1} is geometrically convex. By part (b) of Lemma 6, geometric convexity implies that the conjugate functions to u and u^{-1} are both convex. Thus $\hat{u}$ and $\widehat{u^{-1}}$ are both convex. Because $\widehat{u^{-1}} = \hat{u}^{-1}$, $\hat{u}$ and its inverse must both be increasing convex functions equal to 0 at $y = 0$. Thus, $\hat{u}$ must be a linear function of the form $\hat{u}(y) = py,\ p > 0$. Thus,

$$u(t) = \exp(p \log(t)) = t^p,\ p > 0.$$

Thus, we have shown that (i) implies (iv).

To complete the proof, we need only show that (iii) implies (iv). To establish the result, note that, by Theorem 1, CSSD dominance, using the Π function defined by Definition S1-1, is equivalent to

$$\begin{aligned} \Pi[u^{-1}](u(t)) \geq \Pi[u^{-1}](1) = \int_0^1 u^{-1}(s)\,ds,\ t \in [0,1], \\ \Pi[u](t) \geq \Pi[u](1) = \int_0^1 u(s)\,ds,\ t \in [0,1]. \end{aligned} \tag{S1.9}$$

Expanding the definition of $\Pi[u^{-1}](u(t))$, we see that

$$\Pi[u^{-1}](u(t)) = \frac{1}{u(t)\,t} \int_0^t u^{-1}(s).\,ds. \tag{S1.10}$$

Young's Theorem (see for example Theorem 156 in Hardy, Littlewood, and Polya (1952)) implies that

$$\int_0^t u(s)\,ds + \int_0^{u(t)} u^{-1}(s)\,ds = t\,u(t). \tag{S1.11}$$

Equations (S1.11) and (S1.10) and imply that

$$\Pi[u^{-1}](u(t)) = 1 - \Pi[u](t).$$

Letting $t = 1$ in (S1.11) shows that

$$\int_0^1 u(s)\,ds + \int_0^1 u^{-1}(s)\,ds = 1. \tag{S1.12}$$

Thus, if we let C equal the first integral in (S1.12), we see that the inequalites in expression (S1.9) imply that

$$\Pi[u](t) \geq C \quad \text{and} \quad 1 - \Pi[u](t) \geq 1 - C.$$

Thus, $\Pi[u](t) = C$. This implies that for all $t \in (0,1]$,

$$\frac{1}{C}\frac{1}{t}\int_0^t u(s)\,ds = u(t). \tag{S1.13}$$

Because u is identically equal to the left-hand side of equation (S1.13), and because u is continuous and thus its integral is differentiable, u must be differentiable. Differentiation of equation (S1.13) shows that u must satisfy the differential equation,

$$(1-C)\,u(t) - C\,t\,u'(t) = 0, \quad u(1) = 1.$$

This differential equation has a unique solution, $u(t) = t^{(1-C)/C}$. $\square$

Proof of Theorem 4. From Chan, Proschan, and Sethuraman (1990) we see that

$$u(t) = \int_0^t \phi \circ G^{-1}(s)\, ds, \tag{S1.14}$$

where ϕ is the Radon-Nikodym derivative of G with respect to F. If G and F are a regular pair of distributions, ϕ is absolutely continuous with respect to Lebesgue measure and is given by

$$\phi(x) = \frac{f(x)}{g(x)}.$$

Thus,

$$u'(t) = \phi \circ G^{-1}(t) = \frac{f(G^{-1}(t))}{g(G^{-1}(t))}. \tag{S1.15}$$

First consider (i). Because G is continuous and its support is $[\underline{x}, \bar{x}]$, G strictly increasing and thus G^{-1} is strictly increasing over $[\underline{x}, \bar{x}]$. Hence u increasing if and only if ϕ is increasing, i.e., f/g is increasing.

Now consider (ii), For regularly related distributions, geometric convexity requires that R, defined in Lemma 7, be increasing. Substituting the definitions of u and u' from equations (S1.14) and $u = F \circ G^{-1}$ into R shows that

$$R(t) = \frac{\phi \circ G^{-1}(t)\, t}{F \circ G^{-1}(t)}. \tag{S1.16}$$

Now make the substitution $x = G(t)$. This yields

$$R \circ G(x) = \frac{\phi(x)\, G(x)}{F(x)}.$$

Because G is strictly increasing, $x \hookrightarrow R \circ G(x)$ is increasing if and only if $x \hookrightarrow (\phi(x)\, G(x))/F(x) = (f(x)\, G(x))/(g(x)\, F(x))$ is increasing.

□

Proof of Theorem 5. First consider (i). $F < G$ on some neighborhood of $\underline{x}$, implies that $u(t) < t$ on some neighborhood of 0. The geometric convexity of u implies by Lemma 6 that conjugate function to u, $\hat{u}(y) = \log \circ u \circ \exp(y)$, is an increasing convex function defined over $(-\infty, 0]$. The conjugate function to the identity function $\mathrm{id}(t) = t$ is simply $\widehat{\mathrm{id}}(y) = y$, the identity function. Because conjugation preserves order relations, and because $u(t) > \mathrm{id}(t)$ in a neighborhood of 0, condition (ii) implies that there exists $\underline{y} < 0$, such that $\hat{u}(y) < \widehat{\mathrm{id}}(y)$ when $y < \underline{y}$. Because $\hat{u}(y)$ is convex and $\widehat{\mathrm{id}}(y)$ is linear and because the functions meet at 0, they cannot meet at any other point. Thus, $\hat{u}(y) < \mathrm{id}(y)$ for all $y < 0$. The order-preserving nature of conjugation then ensures that $u(t) < t$, for $t < 1$. The definition of u then implies that $F(x) < G(x),\ x \in (\underline{x}, \bar{x})$, Thus, F strictly stochastically dominates G.

Now consider (ii). $F > G$ on an open neighborhood of $\underline{x}$, implies that $u(t) > t$ on some open neighborhood of 0. Thus, for the same reasons as advanced in the proof of part (i), there exists $\underline{y} > 0$, such that $\hat{u}(y) > \widehat{\mathrm{id}}(y)$ when $y < \underline{y}$. Because $\hat{u}(y)$ is continuous, either (case (a)) $\hat{u}(y) > \widehat{\mathrm{id}}(y),\ y < 0$ or (case (b)) there exists $y^o < y$ such that $\hat{u}(y^o) = \widehat{\mathrm{id}}(y^o)$. In case (a), $\hat{u}(y) > \widehat{\mathrm{id}}(y),\ y < 0$ implies that $u(t) > t,\ t \in (0,1)$. The definition of u then implies that $F(x) > G(x),\ x \in (\underline{x}, \bar{x})$. In case (b), because F strictly geometrically dominates G, $\hat{u}$ is strictly convex. Because, $\widehat{\mathrm{id}}$ is linear, its value can equal the value of $\hat{u}$ at, at most, two points. $\hat{u}(y) > y = \widehat{\mathrm{id}}(y)$ when $y < \underline{y}$ and $\hat{u}(0) = 0 = \widehat{\mathrm{id}}(0)$. Thus, y^o is the unique $y < 0$ such that $\hat{u}(y) = y$. The continuity of $\hat{u}$, the intermediate value theorem, and the fact that $\hat{u}(y) > y$ when $y < \underline{y}$, then imply that $\hat{u}(y) > y^o$ whenever $y < y^o$ and $\hat{u}(y) < y^o$ whenever $0 > y > y^o$, which implies that F crosses G once from above. $\square$

Proof of Theorem 6. We start by proving assertion (i). For any real valued function of a single variable, f, let $D_+ f(x)$ represent the right derivative of f evaluated at x. Note that a convex function has a right derivative at all points on the interior of its domain and that a necessary and sufficient condition for convexity of a function is that its right derivative is increasing.

Define the function $\hat{v} : (-\infty, 0] \to \mathbb{R}$ by

$$\hat{v}(y) = \hat{u}(y) - y, \quad y \leq 0.$$

Where $\hat{u}$ is conjugate function to u defined in Lemma 6. First note that, because u is strictly geometrically convex, $\hat{u}$ is strictly convex and thus $\hat{v}$ is strictly convex. The hypothesis of strict stochastic dominance implies that $u(t) < t, t \in (0,1)$. Because the conjugate transform is order preserving, this implies that $\hat{u}(y) - y < 0$, Thus, $\hat{v}$, which is defined on the non-positive real line, is bounded from above by the x-axis. Thus $\hat{v}$ must be increasing. To see this note that if $\hat{v}$ were decreasing anywhere, it would have a support line with a negative slope. For y sufficiently small this support line would cross the x-axis. Because $\hat{v}$ is convex its graph lies above all of its support lines. Thus, $\hat{v}$ would cross the x-axis. Because, by hypothesis, such crossing is impossible, it must be the case that $\hat{v}$ is increasing.

Using the definition of the conjugate function given in Lemma 6 we can see that

$$\hat{v} \circ \log(t) = \hat{u} \circ \log(t) - \log(t).$$

The definition of the conjugate function implies that $u(t) = \exp \circ \hat{u} \circ \log(t)$. Thus,

$$u(t) = t \exp[\hat{v}(\log(t))]. \tag{S1.17}$$

Because $\hat{u}$ is convex, it has left and right derivatives. Thus, u, being the composition of $\hat{u}$ with the smooth functions, exp and log, has left and right derivatives. We differentiate equation (S1.17) and obtain

$$\begin{aligned} D_+ u(t) = \exp[\hat{v}(\log(t))] + t \exp[\hat{v}(\log(t))] \, D_+ \hat{v}(t) \, \frac{1}{t} = \\ \exp[\hat{v}(\log(t))] + \exp[\hat{v}(\log(t))] \, D_+ \hat{v}(t), \quad t \in (0,1). \end{aligned} \tag{S1.18}$$

Because $\hat{v}$ is increasing and convex, $D_+ \hat{v}$ is positive and increasing. Thus, equation (S1.18) implies that $D_+ u$ is increasing over (0, 1), which implies that u is convex.

Now note that Note that the second assertion in the Theorem immediately follows from (i) and part (i) of Theorem 4. $\square$

Proof of Lemma 12. To prove the assertion for geometric convexity and geometric convexity on average define the map, $\Phi : (-\infty, 0] \to \Re$ by $\Phi(y) = \log[p + (1-p)\, e^y]$. Computing the derivative shows that $\Phi'' > 0$ and thus Φ is strictly convex and it is obvious that it is strictly increasing. Let $u_o = F_o \circ G^{-1}$ and let $u = F \circ G^{-1}$ Then $\hat{u} = \Phi \circ \hat{u}_o$. By assumption, $\hat{u}_o$ is convex. Φ is an increasing convex map, and thus $\Phi \circ \hat{u}_o = \hat{u}$ is strictly convex, i.e., F strictly geometrically dominates G. Exactly the same argument, with U replacing u, works to prove the assertion for geometric convexity on average.

Now consider CSSD. Using Definition S1-1, we can write

$$\Pi[u](t) = \frac{p + (1-p)\,U_o(t)}{p + (1-p)\,u_o(t)} = \frac{\theta(t) + (1-p)\,\Pi[u_o](t)}{\theta(t) + (1-p)}, \quad \theta(t) = \frac{p}{u(t)}.$$

Because, $\Pi\,[u_o] < 1$, the function

$$\theta \hookrightarrow \frac{\theta + (1-p)\,\Pi[u_o](t)}{\theta + (1-p)}$$

is strictly increasing in θ. Moreover, if $t \in (0,1)$, $\theta(t) > \theta(1)$. Thus,

$$\Pi[u](t) = \frac{\theta(t) + (1-p)\,\Pi[u_o](t)}{\theta(t) + (1-p)} > \frac{\theta(1) + (1-p)\,\Pi[u_o](t)}{\theta(1) + (1-p)}, \; t \in (0,1). \tag{S1.19}$$

By the assumption of CSSD dominance of F_o over G, $\Pi[u_o](t) \geq \Pi[u_o](1)$. Thus,

$$\frac{\theta(1) + (1-p)\,\Pi[u_o](t)}{\theta(1) + (1-p)} \geq \frac{\theta(1) + (1-p)\,\Pi[u_o](1)}{\theta(1) + (1-p)} = \Pi\,[u](1). \tag{S1.20}$$

Expressions (S1.19) and (S1.20) imply that, $\Pi[u_o](t) > \Pi[u_o](1)$, $t \in (0,1)$, i.e., F strictly CSSD dominates G. □

Proof of Theorem 7. We first establish (i). First note that if F places positive mass on $\underline{x}$ then because G places no mass on $\underline{x}$, the result is trivially true. So suppose that F places no mass on $\underline{x}$. In this case, $u(0) = U(0) = 0$.. Note that, Π, defined in Definition S1-1, verifies $\Pi[u](1) = U(1)/u(1) = U(1)$ and that F being strictly stochastically dominated implies, by Lemma S1-4, that $U(1) > 1/2$. Thus, CSSD requires that $\Pi[u](t) = U(t)/u(t) > \Pi[u](1) > 1/2$. Thus, there exists $c > 0$, such that $U(t)/u(t) \geq 1/2 + c$. Next, note that expression (11) shows that $u(t) = t\,U'(t) + U(t)$. Hence,

$$U(t) \geq \left(\frac{1}{2} + c\right) u(t) = \left(\frac{1}{2} + c\right) \Big(t\,U'(t) + U(t)\Big).$$

Letting $K = (\frac{1}{2} + c)/(\frac{1}{2} - c)$, algebraic rearrangement shows that there exists $K > 1$ such that

$$K\,U'(t) \leq \frac{U(t)}{t}, \text{for all } t \in (0,1).$$

Because U is differentiable on $(0,1)$ and continuous on $[0,1]$ we can apply the mean value theorem, to show that there exists $\eta(t) \in (0,t)$ such that

$$K\,U'(t) \leq \frac{U(t) - U(0)}{t - 0} = U'(\eta(t)) \leq \sup_{s \in (0,t)} U'(\eta(s)).$$

Thus,

$$\limsup_{t \to 0} K\,U'(t) \leq \limsup_{t \to 0} \frac{U(t)}{t} \leq \limsup_{t \to 0} U'(t).$$

If $U(t)/t$ is bounded on $(0,t)$.

$$\limsup_{t \to 0} U'(t) = L < \infty,$$

and thus, expressions (S1.3) and (S1.3) imply that $KL \leq L$. But, this is not possible because $K > 1$. Thus it must be the case that $\limsup_{t\to 0} U(t)/t = \infty$. Next note that

$$\frac{U(t)}{t} = \frac{\int_0^t \frac{u(s)}{s} s\,ds}{t^2} = \frac{\int_0^t \frac{u(s)}{s} s\,ds}{2\int_0^t s\,ds}. \tag{S1.21}$$

Because $u(s)/s$ is continuous for $s > 0$ and $\limsup_{t\to 0} U(t)/t = \infty$, $u(s)/s$ must be unbounded on a neighborhood of 0. For if $u(s)/s$ were bounded on a neighborhood of 0 there would exist $B < \infty$ such that $u(s)/s < B$ uniformly on this neighborhood, which by Lemma S1-4 implies that $U(t)/t \leq B/2$. As this is not possible given that $\limsup U(t)/t \to \infty$ we conclude that $\limsup u(t)/t = \infty$. Substituting $G(x)$ for t then shows that the assertion in part (i) must hold.

Now consider part (ii). Define $\hat{v}(t) = \hat{u}(t) - t$, where $\hat{u}$ is the conjugate function to u, (defined in Lemma 6). Note that $u(1) = 1$, which implies that $\hat{u}(0) = 0$. The fact that F is strictly stochastically dominated implies that $u(t) > t$, $t \in (0,1)$, thus,

$$\hat{v}(y) \geq 0,\, y \leq 0 \text{ and } \hat{v}(0) = 0. \tag{S1.22}$$

Because $\hat{v}$ is strictly convex, (S1.22) implies that $\hat{v}$ is strictly decreasing. To see this, note that because $\hat{v} \geq 0$ and $\hat{v}(0) = 0$, $\hat{v}$, and $\hat{v}$ is strictly convex and thus monotone on intervals, it is strictly decreasing in neighborhood of 0. Convexity, then implies that it must be strictly decreasing for all $y < 0$. Using equation (S1.17), we can write

$$\frac{u(t)}{t} = \exp[\hat{v}(\log(t))],\ t \in (0,1]. \tag{S1.23}$$

Because, exp and log are strictly increasing functions and $\hat{v}$ is strictly decreasing, $t \to u(t)/t$ is strictly decreasing. Thus, $t \to u(t)/t$ is strictly decreasing. Because $\hat{v}$ is strictly convex and strictly decreasing it is bounded from below by support line with a negative slope. Thus $\lim\limits_{y\to-\infty} \hat{v}(y) = \infty$. Hence, (S1.23) implies that, $\lim\limits_{t\to 0} u(t)/t = \infty$. This implies for the same reasons as given in Lemma S1-6 that, $x \hookrightarrow F(x)/G(x)$ is strictly decreasing. This establishes part (a) of (ii). Part (b) of (ii) follows because geometric dominance implies geometric dominance on average by Lemma 9 part (b) holds. □

Proof of Lemma 13. The distribution of $\tilde{X} + c$, $c > 0$, is $G(x - c)$, $x \in \Re$. Thus, the transform function, u, is given by

$$u(t) = G(G^{-1}(t) - c).$$

Using the inverse function theorem we see that

$$u'(t) = \frac{G'(G^{-1}(t) - c)}{G'(G^{-1}(t))} = \frac{g(G^{-1}(t) - c)}{g(G^{-1}(t))}.$$

Thus, the function $R = (tu')/u$ defined in Lemma 7 is given by

$$R(t) = \frac{g(G^{-1}(t) - c)}{G(G^{-1}(t) - c)} \frac{t}{g(G^{-1}(t))}.$$

Strict geometric dominance is equivalent to R being strictly increasing. Make the substitution $x = G^{-1}(t)$. This yields

$$R \circ G(x) = \frac{g(x - c)}{G(x - c)} \frac{G(x)}{g(x)} = \frac{r(x - c)}{r(x)}.$$

The ratio $r(x - c)/r(x)$ is (strictly) increasing for every choice of $c > 0$ if and only if

$$\forall c > 0, x \hookrightarrow \log \circ r(x - c) - \log \circ r(x) \text{ is (strictly) increasing over } \Re. \tag{S1.24}$$

Because $\log \circ r$ is continuous, Condition (S1.24) holds if and only if $\log \circ r$ is concave (Theorems 7.3.3 and 7.3.4: Kuczma, 2000). □

Proof of Lemma 14. The distribution of $s\tilde{X}$, $s > 1$, is $G(x/s)$, $x \in \Re^+$. Let $\theta = 1/s < 1$ Thus, the transform function, u, is given by

$$u(t) = G(\theta\, G^{-1}(t)).$$

Using the inverse function theorem, we see that

$$u'(t) = \frac{\theta\, G'(\theta\, G^{-1}(t))}{G'(G^{-1}(t))} = \frac{\theta\, g(\theta\, G^{-1}(t))}{g(G^{-1}(t))}.$$

Thus, the function $R = (tu')/u$ defined in Lemma 7 is given by

$$R(t) = \frac{\theta\, g(\theta\, G^{-1}(t))}{G(\theta\, G^{-1}(t))}\, \frac{t}{g(G^{-1}(t))}.$$

(Strict) geometric dominance is equivalent to R being (strictly) increasing. Make the substitution $x = G^{-1}(t)$, this yields

$$R \circ G(x) = \frac{\theta g(\theta\, x)}{G(\theta\, x)}\, \frac{G(x)}{g(x)} = \frac{\theta\, r(\theta\, x)}{r(x)}.$$

The ratio $\theta\, r(\theta\, x)/r(x)$ is (strictly) increasing for every choice of $\theta \in (0,1)$ if and only if

$$\forall\theta \in (0,1),\, x \hookrightarrow \log \circ r(\theta\, x) - \log \circ r(x) \text{ is (strictly) increasing over } \Re^+. \tag{S1.25}$$

Next, let $\hat{r}$ represent the conjugate function to r defined in Lemma 6. Let $y = \log(x)$ and let $\beta = -\log(\theta)$ and note that $\log \circ r = \hat{r} \circ \log$. Thus, condition (S1.25) is equivalent to the condition that

$$\forall\beta > 0,\, y \hookrightarrow \hat{r}(y - \beta) - \hat{r}(y) \text{ is (strictly) increasing over } (-\infty, 0). \tag{S1.26}$$

By exactly the same argument given in Lemma 13, condition (S1.26) is equivalent to $\hat{r}$ being (strictly) concave and $\hat{r}$ being (strictly) concave is equivalent to r being (strictly) geometrically concave. □

Lemma S1-7. *If G is a probability distribution supported by $(-\infty, \infty)$ with a differentiable density, g, and logconcave reversed hazard rate, r, then r is decreasing.*

Proof. If r is the reversed hazard rate of G then

$$G(x) = \exp\left[-\int_x^\infty r(u)\, du\right].$$

The fact that $\lim_{x \to -\infty} G(x) = 0$ implies that

$$\lim_{x \to -\infty} \int_x^\infty r(u)\, du = \infty. \tag{S1.27}$$

Because $G(x) > 0$, for all $x \in (-\infty, \infty)$,

$$\int_x^\infty r(u)\, du < \infty. \tag{S1.28}$$

If r is logconcave then $\log \circ r$ is concave and thus is majorized by all of its support lines. Because $\log \circ r$ is concave, it is right differentiable. Suppose that for some $w \in \Re$, $D_+ r(w) = r_o > 0$, then, $u \hookrightarrow r(w) + r_o\,(u - w)$ is a support line for r and thus majorization implies that

$$\text{for all } u \in \Re, \quad \log \circ r(u) \le r(w) + r_o\,(u - w), \quad r_o > 0. \tag{S1.29}$$

Exponentiating both sides of equation (S1.29) and then integrating u over (x,w) shows that for $x<w$,

$$\int_x^w r(u)\,du \le e^{r(w)} \int_x^\infty \exp[r_o(u-w)]\,du = e^{r(w)}\,\frac{1-e^{r_o(x-w)}}{r_o} \le \frac{e^{r(w)}}{r_o}.$$

Hence, for $x<w$

$$\int_x^\infty r(u)\,du = \int_x^w r(u)\,du + \int_w^\infty r(u)\,du < \frac{e^{r(w)}}{r_o} + \int_w^\infty r(u)\,du,$$

which implies that

$$\lim_{x\to-\infty}\int_x^\infty r(u)\,du \le \frac{e^{r(w)}}{r_o} + \int_w^\infty r(u)\,du < \infty,$$

contradicting expression (S1.27). Thus, it must be the case that, for all $x \in \Re$, $D_+r(x) \le 0$, which implies that r is decreasing. □

Lemma S1-8. *If r is the reversed hazard rate for a probability distribution, G, supported by $[0,\infty)$ and r is geometrically concave then r is decreasing.*

Proof. First, note that, by the same argument as given in Lemma S1-7, the fact that r defines a distribution function implies that

$$\int_x^\infty r(u)\,du < \infty,\ x>0, \tag{S1.30}$$

$$\lim_{x\to 0}\int_x^\infty r(u)\,du = \infty. \tag{S1.31}$$

Next, note that by definition (Lemma 5), if r is geometrically concave then its conjugate $\hat{r} : \Re \to \Re$ is concave. Thus, by concavity, the set of points at which $\hat{r}$ is strictly increasing is either empty or an interval $(-\infty, c)$. Since conjugation preserves order, the set of points where r is strictly increasing is either empty or the interval $(\exp[-\infty], \exp[c]) = [0, \exp[c])$. Thus if r were strictly increasing anywhere it would be increasing over some interval of the form $[0,C)$, where $C = \exp[c]$. Thus, becasue r is increasing over this interval,

$$\int_0^C r(s)\,ds \le r(C)\,C.$$

Hence, for all $x<C$,

$$\int_x^\infty r(s)\,ds \le r(C)\,C + \int_C^\infty r(s)\,ds. \tag{S1.32}$$

Thus, expressions (S1.30) and (S1.32) imply that

$$\lim_{x\to 0}\int_x^\infty r(s)\,ds \le r(C)\,C + \int_C^\infty r(s)\,ds < \infty.$$

contradicting expression (S1.31). Thus, r cannot be strictly increasing anywhere and thence r is decreasing. □

Lemma S1-9. *Suppose that G and probability distribution function, supported by $(-\infty,\infty)$, with a continuously differentiable density, g, and logconcave reversed hazard rate function, r, then g, is logconcave.*

Proof. The reversed hazard rate, $r = f/F$ being logconcave implies that $r'/r = f'/f - r$ is decreasing. By Lemma S1-7, r is decreasing. Thus, r'/r can only be decreasing if f'/f be decreasing, i.e., the probability density is logconcave □

Example S1.1 (Example of a distribution function supported by $[0,\infty)$ with a strictly geometrically concave reversed hazard rate and a strictly logconvex PDF)**.** Consider the log-logistic distribution, G, with shape parameter $\beta = 1$ and size parameter $\alpha = 1$ (See Table 1). Simple calculations show that the reversed hazard rate for this distribution, r, and its log derivative, g'/g, are given by

$$r(x) = \frac{1}{x^2+x}, \qquad \frac{g'(x)}{g(x)} = -\frac{2}{1+x}.$$

Because g'/g is strictly increasing the probability density is not only not logconcave, it is strictly logconvex. However, simple calculations reveal that

$$\frac{xr'(x)}{r(x)} = -\left(\frac{x}{1+x}+1\right).$$

Thus $x \hookrightarrow (xr'(x))/r(x)$ is strictly decreasing and thus the necessary and sufficient condition for geometric concavity given in Lemma 7 is satisfied and, hence, the reversed hazard rate of G is geometrically concave.

Example S1.2 (Example of a distribution function supported by $[0,\infty)$ with a strictly logconcave density whose reversed hazard rate which is not geometrically concave)**.** Consider the distribution function, G, defined by

$$G(x) = \exp\left[-\int_x^\infty r(s)\,ds\right], \text{where } r(s) = \frac{e^{(1-s)s}}{s}.$$

the reversed hazard rate for this distribution, r, and its log derivative, g'/g are given by

$$r(x) = \frac{e^{(1-x)x}}{x}, \qquad \frac{g'(x)}{g(x)} = \frac{-1+e^{x-x^2}+x-2x^2}{x}. \tag{S1.33}$$

Computations show that

$$\frac{d}{dx}\left(\frac{g'(x)}{g(x)}\right) = \frac{e^{x-x^2}\left(\Psi_1(x)+\Psi_2(x)\right)}{x^2},$$
$$\Psi_1(x) = e^{x^2-x}\left(1-2x^2\right),$$
$$\Psi_2(x) = -1+x-2x^2.$$

Note that

$$\operatorname{sgn}\left[\left(\frac{g'(x)}{g(x)}\right)\right] = \operatorname{sgn}[\Psi_1(x)+\Psi_2(x)].$$

Thus, establishing logconcavity of the density function is equivalent to showing that $\Psi_1(x) + \Psi_2(x) < 0$, for $x > 0$. Note that $\Psi_2 < 0$ for all positive x and that $\Psi_1 < 0$ whenever $x > \frac{1}{\sqrt{2}}$. Thus, if $\Psi_1 + \Psi_2$ were ever positive it would be positive over $[0, \frac{1}{\sqrt{2}}]$. Over this range, both Ψ_1 and Ψ_2 are strictly concave and thus their sum is strictly concave. Evaluating the derivative of $\Psi_1 + \Psi_2$ at $x = 0$ reveals that $\Psi_1'(0) + \Psi_2'(0) = 0$. Because $\Psi_1 + \Psi_2$ is strictly concave and continuously differentiable, this implies that $\Psi_1 + \Psi_2$ attains a strict maximum at $x = 0$. Thus, $\Psi_1(x) + \Psi_2(x) < \Psi_1(0) + \Psi_2(0)$, $x > 0$. But, $\Psi_1(0) + \Psi_2(0) = 0$, and thus $\Psi_1(x) + \Psi_2(x) < 0$, $x > 0$, and the result is established, i.e., G has a logconcave probability density.

Now consider the reversed hazard rate of G. Using the expression for r given by (S1.33), we see that

$$\frac{xr'(x)}{r(x)} = -1 + x - 2x^2,$$

Thus, $x \hookrightarrow (x\,r'(x))/r(x)$ is not monotone, rather it is increasing for $x < 1/4$ and decreasing for $x > 1/4$. Thus, the necessary and sufficient condition for geometric concavity given in Lemma 7 fails to be satisfied and, hence, the reversed hazard rate of G is not geometrically concave.

S2 Example of the intransitivity of selection dominance

Define the functions: $u_1 : [0,1] \to [0,1]$, $u_1 : [0,1] \to [0,1]$ as follows. Let

$$u_o(t) = \begin{cases} \frac{1}{2}\frac{t}{\eta_o} & \text{if } t \in [0, \eta_o) \\ \frac{1}{2} & \text{if } t \in [\eta_o, 1-\eta_o) \\ \frac{1}{2} + \frac{1}{2}\frac{t-(1-\eta_0)}{\eta_o} & \text{if } t \in [\eta_o, 1] \end{cases},$$

where $\eta_o = 3/50$. Let

$$u_1(t) = p_o t + (1-p_o)\, u_o(t) \tag{S2.1}$$

$$u_2(t) = \frac{(t+1)\log(t+1) - c_o t}{2\log(2) - c_o}, \tag{S2.2}$$

where $c_o = 9/10$ and $p_o = 1/10$. It is easy to verify that u_1 and u_2 are an admissible functions. Thus, these functions define an admissible collection of distributions, F, G, and H over the unit interval:

$$H(x) = x, \quad G(x) = u_2 \circ H(x), \quad F(x) = u_1 \circ G(x).$$

These distributions, as well as their associated selection-dominance functions, Π, defined by Definition (S1-1), are graphed in Figure S2. Panels B and C of Figure S2 verify that u_1 and u_2 satisfy the CSSD condition Lemma S1-1. Thus, F CSSD dominates G and G CSSD dominates H. Because $F(x) = u_1 \circ u_2 \circ H(x)$, for F to selection dominate H it is necessary for $u_1 \circ u_2$ to satisfy the CSSD condition given in expression (12). This condition requires that $\Pi[u_1 \circ u_2]$ have a minimum value at $t = 1$. As Panel D shows, this is not the case. Thus, F does not dominate H by CSSD. Hence, the CSSD relation is not transitive.

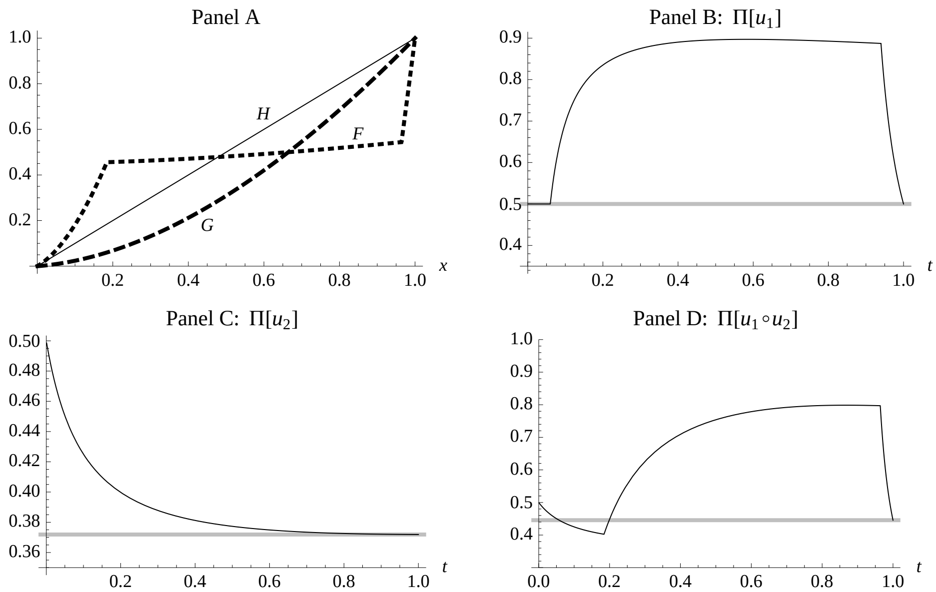


Figure 1: Counterexample to transitivity of selection dominance. Panel A plots the distribution functions, F G and H. Panel B plots the function, $\Pi(u_1)$, (defined by Definition S1-1) used to test the CSSD dominance of F over G, Panel C plots the function, $\Pi(u_2)$, used to test the CSSD dominance of G over H. Panel D plots the function, $\Pi(u_1 \circ u_2)$, used to test the CSSD dominance of F over H.

S3 Distributions

Preliminary remarks

We will make use of the following theorem frequently and thus, to minimize repetition, we provide it here and, in the sequel, simply refer to it as MLHR.

Theorem 1 (Monotone LHospital Rule **MLHR** (Proposition 1.1 in Pinelis (2002)))**.** *Let $-\infty \leq a < b \leq \infty$. Let ϕ and ν be real-valued differentiable functions defined on (a,b) and that Suppose that either $\phi(a+) = \nu(a+) = 0$ or $\phi(b-) = \nu(b-) = 0$, and that ν' does not change sign over the interval (a,b), then*

$$x \hookrightarrow \frac{\phi'(x)}{\nu'(x)} \text{ is (strictly) increasing} \Rightarrow x \hookrightarrow \frac{\phi(x)}{\nu(x)} \text{ is (strictly) increasing.}$$

We try to distinguish the specific properties of the given distribution being analyzed from general properties of the distributions and their associated densities and reversed hazard rate. For the short derivations, we do this by labeling the densities, distributions and hazard rates, with a subscript indicating the specific distribution being analyzed. For the longer derivations, e.g., Gompertz and Gamma distributions, this is impractical because it leads to very bulky expressions. Therefore, in sections devoted to these distributions, we simply state that all references to distributions, densities, reversed hazard rates, in the given section refer to the specific densities, distributions, and hazard rates of the distribution being analyzed in that section. Note also that we are dealing with textbook distributions. Only one of the distributions considered, the Laplace distribution, lacks smooth derivatives of all orders. And even the Laplace is smooth at all points in its support save one. Thus, we will frequently use continuous differentiability in the proofs without explicitly invoking continuous differentiability is our arguments.

S3.1 Normal

The Mill's ratio for the Normal distribution is log convex (see, e.g., Sampford (1953)). This implies that the hazard rate, h_{N} of the Normal distribution is log concave. The symmetry of the normal distribution implies that, the hazard rate and reversed hazard rate, r_N, are related by $h_{\text{N}}(-x) = r_{\text{N}}(x)$, $x \in \Re$. Note that $\log \circ h_{\text{N}}(x)$ is concave if and only if $\log \circ h_{\text{N}}(-x) = r_{\text{N}}(x)$ is concave. Thus r_N is logconcave. Geometric dominance follows from Lemma 13.

S3.2 Logistic

Assume without loss of generality that $s = 1$ and $\mu = 0$. Under this assumption, the reversed hazard rate for the logistic is given by

$$r_{\text{Logistic}}(x) = \frac{1}{1+e^{-x}}, x \in \Re.$$

Thus,

$$\log \circ r_{\text{Logistic}}(x) = -\log\left(1+e^{x}\right).$$

Next note that

$$(\log \circ r_{\text{Logistic}}(x))'' = -\frac{e^{x}}{(1+e^{x})^2} < 0.$$

Thus r_{Logistic} is logconcave and geometric dominance thus follows from Lemma 13.

S3.3 Laplace

Let G_{Laplace} and g_{Laplace} represent the distribution and density of the standard ($s = 1$, $\mu = 0$) Laplace distribution. First, we show that geometric dominance fails in general and then show that geometric dominance on average holds. Note that the reversed hazard rate for the Laplace, r_{Laplace} is given by $r_{\text{Laplace}} = (2\exp(\max[x,0]-1))^{-1}$. It is easy to verify that, restricted to $x > 0$, r_{Laplace} is log convex and thus by Lemma 13, G is not geometrically dominated by all of its upshifts.

Let F_{Laplace} represent a μ-upshift of G_{Laplace}. We verify geometric convexity on average using the conditions given in Lemma S1-5 to show that the function

$$\mathscr{R}_{\text{Laplace}}(x) = \frac{\int_{-\infty}^{x} g_{\text{Laplace}}(z-\mu)\, G_{\text{Laplace}}(z)\, dz}{\int_{-\infty}^{x} g_{\text{Laplace}}(z)\, G_{\text{Laplace}}(z-\mu)\, dz}.$$

is increasing. After considerable, tedious, but quite standard, calculus calculations, one can express $\mathscr{R}_{\text{Laplace}}$ in the following somewhat compact form. Let $w = e^x$, and let $m = e^{\mu}$, then

$$\mathscr{R}_{\text{Laplace}}(w) = \begin{cases} 1 & \text{if } w \le 1, \\ T_2(w,m) & \text{if } w \in (1,\mu], \\ T_3(w,m) & \text{if } w > \mu, \end{cases} \tag{S3.1}$$

$$T_2(w,m) = \frac{4w-2}{1+2\log(w)} - 1, \tag{S3.2}$$

$$T_3(w,m) = \frac{w^2\,(8m-4-2\,\mathrm{Log}(m)) - 4m^2 w + m^2}{w^2\,(4+2\log(m)) - 4mw + m^2}. \tag{S3.3}$$

Note that $\mathscr{R}_{\text{Laplace}}$ (where $\mathscr{R}$ is defined by Definition S1-3) is continuous and thus, to verify that it is increasing, one needs only verify that it is increasing on each leg of its definition. Since $\mathscr{R}_{\text{Laplace}}$ is constant on the first leg, the assertion is verified for this leg. Now consider the second leg where $w \in (1,\mu]$. Note that

$$\frac{\partial T_2}{\partial w} = \frac{4\,(2w\log(w) - (w-1))}{w\,(1+2\,\mathrm{Log}(w))^2}.$$

$w \hookrightarrow w\log(w)$ is strictly convex for $w \ge 1$. It support line at 1 is $w-1$. Strictly convex functions exceed their support lines. Thus, $\partial T_2/\partial w > 0$ and hence, T_2 is strictly increasing in y.

Now consider the third leg where $y > \mu$. Note that

$$\frac{\partial T_3}{\partial w} = \frac{4m\,S(w,m)}{\left((2w-m)^2 + 2w^2\,\mathrm{Log}(m)\right)^2}, \tag{S3.4}$$

$$S(w,m) = 2w^2\,Q(m) + 2wm\,(2\,(m-1) - \log(m)) - (m-1)\,m^2, \tag{S3.5}$$

$$Q(m) = 2\,((1+m)\log(m) - 2\,(m-1)). \tag{S3.6}$$

First note that differentiation shows that Q is strictly convex for $m > 1$. Next note that $Q'(1) = Q(1) = 0$. This fact combined with the strict convexity of Q implies that $Q(m) > 0$, for all $m > 1$.

Q is positive and S is strictly convex in y. Thus, to show that S is positive on the third leg, we need only show that evaluated at $y = m$, S is nonnegative and increasing.

$$S(y=m,m) = m^2\,(2m\log(m) - (m-1)), \tag{S3.7}$$

$$\frac{\partial S}{\partial w}(y=m,m) = 2m\,(\log(m) + 2\,(m\log(m) - (m-1))). \tag{S3.8}$$

Because $m > 1$ on the third leg, by the same argument as used in the proof for branch two, $m\log(m) - (m-1) > 0$. Thus, the right-hand sides of (S3.7) and (S3.8) are both positive and hence, the strict convexity of S in y implies that $S > 0$. Inspection of (S3.4) shows that this implies that $\frac{\partial T_3}{\partial w} > 0$, and thus on third leg, $\mathscr{R}_{\text{Laplace}}$ is strictly increasing. Hence, we have established that $\mathscr{R}_{\text{Laplace}}$ is increasing in y. Because $y = e^x$, $\mathscr{R}_{\text{Laplace}}$ is increasing in x. Thus, by Lemma S1-5, the translated Laplace distribution always geometrically dominates on average the original distribution but does not dominate it geometrically. Next note that geometric dominance on average implies that $t \hookrightarrow \frac{U'(t)t}{U(t)}$ is increasing. The proof also showed that $t \hookrightarrow \frac{U'(t)t}{U(t)}$ is strictly increasing for t sufficently large, thus $t \hookrightarrow \frac{U'(t)t}{U(t)}$ attains a strict maximum at $t = 1$, which implies, by equations (12) and (18), that the upshifted Laplace strictly CSSD dominates the orignal distribution.

S3.4 Gumbel

Selection equivalence is quite easy to show using many of the tests developed in the manuscript. Perhaps the most transparent demonstration follows from noting that if G is standard $s = 1, \mu = 0$, Gumbel, and F is a c-upshift of G then, the transform function, u, is given by $u(t) = t^C$, where $C = e^c$. The function $t \hookrightarrow t^C$ is clearly geometrically linear. Thus, the c-upshift is geometrically equivalent to the original distribution and hence, by Lemma 10, the distributions are selection equivalent.

S3.5 Gamma

In order to make the notation more manageable. Define $\theta = 1/s$. In this section let G represent a standard unit scale Gamma distribution with shape parameter α, i.e.,

$$G(x) = \frac{1}{\Gamma(\alpha)} \int_0^x z^{1-\alpha} e^{-z} dz = \frac{\gamma(\alpha, x)}{\Gamma(\alpha)},$$

where Γ represents the Gamma function and γ represents the lower incomplete Gamma function.

Let F represent an upscaled Gamma distribution with the same shape parameter α but with scale parameter $1/\theta$, $\theta \in (0,1)$. Thus the distribution of F is given by $G(\theta x)$, $x > 0$, and the reversed hazard rates of F and G are given by

$$r_F(x) = \theta\, r(\theta x); \quad r_G(x) = r(x), \quad x > 0.$$

Note that for $x > 0$, the Gamma distribution, its density, and reversed hazard rate are continuously differentiable. We will use this fact without comment frequently in the sequel. Geometric dominance follows from a series of results.

Result 5.1. The reversed hazard rate, r, for a unit scale Gamma distribution is defined by

$$r(x) = \frac{e^{-x} x^{\alpha-1}}{\gamma(\alpha, x)}, \quad x > 0.$$

The reversed hazard rate for the Gamma, r, has the following basic properties:
(a) r is continuously differentiable on $[0, \infty)$.
(b) r is strictly decreasing.

(c) $\lim_{x\to 0} r(x) = \infty$.
(d) $\lim_{x\to\infty} r(x) = 0$.

Proof. Part (a) follows from the infinite differentiability of the Gamma distribution. Parts (c) and (d) follow from straightforward calculations, and part (b) follows from the log concavity of the Gamma distribution function (Baricz, 2010). $\square$

Result 5.2. The reversed hazard rate of for the Gamma distribution is geometrically concave on a sufficiently small neighborhood of 0.

Proof. Simple but tedious manipulations show that

$$\frac{x\,r'(x)}{r(x)} = -1 + \left(\frac{e^{-x}x^{\alpha}}{\gamma(\alpha,x)}\right)^2 \left(\frac{\gamma(\alpha,x)}{e^{-x}x^{\alpha}} - \frac{\gamma(\alpha,x)}{e^{-x}x^{\alpha+1}}\right). \tag{S3.9}$$

Applying L'Hopital's Rule to the terms of (S3.9) shows that

$$\lim_{x\to 0} \frac{e^{-x}x^{\alpha}}{\gamma(\alpha,x)} = \alpha, \tag{S3.10}$$

$$\lim_{x\to 0} \frac{\gamma(\alpha,x)}{e^{-x}x^{\alpha}} = \frac{1}{\alpha}, \tag{S3.11}$$

$$\lim_{x\to 0} \frac{\gamma(\alpha,x)}{e^{-x}x^{\alpha+1}} = \frac{1}{1+\alpha}. \tag{S3.12}$$

Therefore

$$\lim_{x\to 0} \frac{x\,r'(x)}{r(x)} = -\frac{1}{1+\alpha} < 0.$$

Thus, by Lemma 7, r and satisfies, in some neighborhood of 0, the conditions for geometric concavity $\square$

Result 5.3.
(a) $\lim_{x\to 0} \frac{r_F(x)}{r_G(x)} = \lim_{x\to 0} \frac{\theta\, r(\theta x)}{r(x)} = 1$ and
(b) $\lim_{x\to\infty} \frac{r_F(x)}{r_G(x)} = \lim_{x\to\infty} \frac{\theta\, r(\theta x)}{r(x)} = \infty$.

Proof. First consider part (a). An application of L'Hospital rule shows that

$$\lim_{x\to 0} x\,r_F(x) = \lim_{x\to 0} x\,r_G(x) = \alpha > 0.$$

Thus,

$$\lim_{x\to 0} \frac{r_F(x)}{r_G(x)} = \lim_{x\to 0} \frac{x\,r_F(x)}{x\,r_G(x)} = 1.$$

Now consider part (b). As $x \to \infty$, $F(x) \to 1$ and $G(x) \to 1$. Thus, the limiting behavior of the ratio will be determined by the ratio f/g which clearly approaches infinity at $x \to \infty$. $\square$

Define,

$$\rho(x) = \begin{cases} \frac{r_F(x)}{r_G(x)} & \text{if } x > 0 \\ 1 & \text{if } x = 0. \end{cases} \tag{S3.13}$$

Result 5.4. On a sufficiently small neighborhood of 0, $\rho'(x) > 0$.

Proof. Note from the definition of ρ, and the fact that ρ is positive, ρ' will be positive if and only if

$$\frac{r_F'(x)}{r_F(x)} - \frac{r_G'(x)}{r_G(x)} = \frac{\theta\, r'(\theta x)}{r(\theta x)} - \frac{r'(x)}{r(x)} > 0.$$

For $x > 0$, this condition will be satisfied if and only if

$$\frac{x\theta\, r'(\theta x)}{r(\theta x)} - \frac{x r'(x)}{r(x)} > 0. \tag{S3.14}$$

Because, $\theta \in (0,1)$, $\theta x < x$. Thus Result 5.2 shows that condition (S3.14) is satisfied on a sufficiently small neighborhood of 0. □

Note that, Result 5.3, Result 5.2, and the continuity and differentiability properties of the Gamma distribution, imply the following result

Result 5.5. The ρ function defined in equation (S3.13) verifies the following conditions:
(a) ρ is continuous on $[0,\infty)$.
(b) ρ is continuously differentiable on $(0,\infty)$.
(c) There exits, $\varepsilon > 0$, such that, for all $x \in (0,\varepsilon)$
 (i) $\rho'(x) > 0$.
 (ii) $\rho(x) > 1$.
(d) $\lim_{x\to\infty} \rho(x) = \infty$.

Define

$$\zeta(x) = \frac{f'(x)}{f(x)} - \frac{g'(x)}{g(x)}, \quad x > 0.$$

Explicit calculation shows that

$$\zeta(x) = (1-\theta) > 0. \tag{S3.15}$$

The quotient rule for differentiation shows that ρ verifies the equation

$$\rho'(x) = \rho(x)\Big(\zeta(x) - (r_F(x) - r_G(x))\Big), \quad x > 0. \tag{S3.16}$$

Equations (S3.16) and (S3.15) thus imply that

$$\rho'(x) = \rho(x)\Big((1-\theta) - (r_F(x) - r_G(x))\Big), \quad x > 0. \tag{S3.17}$$

Result 5.6. For all $x > 0$, $\rho(x) > 1$.

Proof. Suppose, to obtain a contradiction, that the result is false. Let $x^o = \inf\{x > 0 : \rho(x) \leq 1\}$. By the continuity of ρ, $\rho(x^o) = 1$. By the definition of x^o, for all $x \in (0,x^o)$, $\rho(x) > 1$. Thus $\rho(x^0) = \min\{\rho(x) : x \in (0,x^o]\}$. However, because $\rho(x^o) = 1$, $r_F(x) - r_G(x) = 0$ and thus by equation (S3.17), in a neighborhood of x^o, $\rho'(x) > 0$. Thus, ρ is increasing to the left of x^o and thus $\rho(x^o)$ cannot be the minimum value of ρ over $(0,x^o]$. □

Result 5.7. For all $x \geq 0$, $\rho'(x) > 0$.

Proof. Let $B = \{x > 0 : \rho'(x) \le 0\}$. By Results 5.5.d and 5.5.c.i, B is bounded and its infimum exceeds 0. By the continuity of ρ', B is open and, if it is not empty, it is thus a countable collection of disjoint open intervals. Select any of these open intervals, say (x_1, x_2), where $0 < x_1 < x_2 < \infty$. By the continuity of ρ', $\rho'(x_1) = \rho'(x_2) = 0$. Because, $\rho' \le 0$ on (x_1, x_2),

$$\rho(x_1) \ge \rho(x_2). \tag{S3.18}$$

Because, $\rho'(x_1) = \rho'(x_2) = 0$, equation (S3.17), and Result 5.6 imply that

$$0 < r_F(x_1) - r_G(x_1) = r_F(x_2) - r_G(x_2). \tag{S3.19}$$

By Result 5.1.b, and the fact that $x_1 < x_2$, $r_G(x_1) > r_G(x_2)$. This fact, and equation (S3.19) imply that

$$\rho(x_1) - 1 = \frac{r_F(x_1) - r_G(x_1)}{r_G(x_1)} < \frac{r_F(x_2) - r_G(x_2)}{r_G(x_2)} = \rho(x_2) - 1, \tag{S3.20}$$

contradicting expression (S3.18). Thus, B is empty and hence, for all $x > 0$, $\rho' > 0$. Thus, ρ is strictly increasing and hence the condition, given in Theorem 4, the geometric dominance of the upscaled distribution, F, is verified. □

S3.6 Generalized Exponential

The reversed hazard rate for a standard ($s = 1$) Generalized Exponential with shape parameter b is given by

$$r_{\text{GExp}}(x) = b\,\frac{e^{-x}}{1 - e^{-x}}.$$

Applying the differential test for geometric concavity given in Lemma 7, yields

$$R_{\text{GExp}}(x) = \frac{x\, r'_{\text{GExp}}(x)}{r_{\text{GExp}}(x)} = \frac{x\, e^x}{1 - e^x}.$$

Differentiating R_{GExp} yields

$$R'_{\text{GExp}}(x) = \frac{e^x}{(1 - e^x)^2}\,(1 + x - e^x).$$

$1 + x$ is the support line for e^x at $x = 1$. Because e^x is convex, e^x lies above the support line and thus $1 + x - e^x < 0$ for $x \ne 1$ and $1 + x - e^x = 0$ at $x = 1$. Hence, R is strictly decreasing, implying that r_{GExp} is strictly geometrically concave. By Lemma 14 this implies that G is strictly dominated by all of its upscalings.

S3.7 Weibull

Let $a_F > a_G$ and suppose that F is distributed Weibull with size parameter a_F and shape parameter λ and that G is distributed Weibull with size parameter a_G and shape parameter λ. Then a simple direct calculation shows the transform function, u associated with F and G is given by

$$u_{\text{Weibull}}(t) = F \circ G^{-1}(t) = 1 - (1 - t)^r, \quad q = \left(\frac{a_G}{a_F}\right)^{\lambda} < 1, \quad t \in [0, 1].$$

Next note that

$$R_{\text{Weibull}} = \frac{t\,u'_{\text{Weibull}}(t)}{u_{\text{Weibull}}(t)} = \frac{q\,(1-t)^{q-1}\,t}{1-(1-t)^q}.$$

Both the numerator and denominator of the above expression vanish at $t = 1$ and the derivative of denominator never changes sign. Thus, we can apply MLHR. The ratio of the derivatives of the numerator and denominator in the above expression is given by

$$1+(1-q)\,\frac{t}{1-t}$$

and this expression is clearly strictly increasing in t for $q < 1$. Thus, u is strictly geometrically convex, and thus F strictly geometrically dominates G.

S3.8 Pareto Distribution

Let $a_F > a_G > 1$ and suppose that F is distributed with size parameter α_F and shape parameter x_m and that G is distributed Pareto with size parameter α_G and shape parameter x_m. Suppose, without loss of generality that $x_m = 1$. Then a simple direct calculation shows the transform function, u associated with F and G is given by

$$u_{\text{Pareto}}(t) = F \circ G^{-1}(t) = 1-(1-t)^q, \quad q = \frac{\alpha_F\,(\alpha_G - 1)}{\alpha_G\,(\alpha_F - 1)} < 1, \quad t \in [0,1].$$

This is exactly the same transform function derived for the Weibull case above. Thus, for the same reasons as for the Weibull, u_{Pareto} is geometrically strictly convex that thus F strictly geometrically dominates G.

S3.9 Kumaraswamy

The reversed hazard rate for the Kumaraswamy with parameters α and b is given by

$$\frac{b\,(1-x^{\alpha})^{b-1}}{1-(1-x^{\alpha})^b}.$$

We can rewrite this expression as

$$\chi(\alpha\,\log(x)), \quad \chi(y) = \frac{b\,e^y\,(1-e^y)^{b-1}}{1-(1-e^y)^b},\ y < 0.$$

Suppose that $\alpha_F > \alpha_G$ the ratio of reversed hazard rates can be expressed as

$$\frac{\chi\,(\alpha_F\,\log(x))}{\chi\,(\alpha_G\,\log(x))}.$$

Because log is a strictly increasing function. This ratio will (strictly) increase if and only if $y \to \chi(\alpha_F\,y)/\chi(\alpha_G\,y)$ is (strictly) increasing over $(-\infty,0)$. By an argument very similar to the argument used to prove Lemma 14, this ratio will be increasing if and only if χ is geometrically convex. Geometric convexity thus depends on the monotonicity properties of

$$R_{\chi}(y) = \frac{y\,\chi'(y)}{\chi(y)}.$$

If R_χ is increasing the stochastically dominant distribution will be geometrically dominant. Similarly, the geometric dominance of the stochastically dominated distribution is equivalent to R_χ being decreasing; stochastic equivalence is equivalent to R_χ being constant. Next, note that

$$R_\chi(y) = \Lambda\left(e^y\right) M(y), \tag{S3.21}$$

$$M(y) = \frac{e^y(-y)}{1-e^y}, \quad y \leq 0, \tag{S3.22}$$

$$\Lambda(z) = \frac{(1-z)^b - (1-bz)}{\left(1-(1-z)^b\right) z}, \quad z \in (0,1). \tag{S3.23}$$

Inspection and an application of MLHR to the right-hand side of equation (S3.22) shows that

$$M \text{ is positive and strictly increasing.} \tag{S3.24}$$

The geometric convexity properties of Kumaraswamy all follow from the following result.

Result 9.1. If $b > 1$, then Λ is positive and strictly increasing; If $b < 1$, Λ is negative and strictly decreasing; If $b = 1$, Λ is identically 0.

Proof of Result. First note that $1 - bz$ is the support line for $(1-z)^b$ at 0. If $b > 1$ then $(1-z)^b)$ is strictly convex in z. Thus, $(1-z)^b > 1 - bz$. If $b < 1$ then $(1-z)^b$ is strictly concave in z. Thus, $(1-z)^b < 1 - bz$. Since the denominator in the expression defining Λ is clearly positive, these observations imply that if $b > 1$, $\Lambda > 0$ and, if $b < 1$, $\Lambda > 0$. Now consider monotonicity. Evaluated at $z = 0$, both the numerator and denominator of Λ vanish and the derivative of the denominator in the expression defining Λ never changes sign. Thus, we can determine the direction of monotonicity using the MLHR. The ratio of the derivatives of the numerator and denominator in the expression defining Λ is given by

$$\text{DRatioKumar} = b\,(1+\lambda(z))^{-1}, \tag{S3.25}$$

$$\lambda(z) = \frac{(1-z)^{b-1} z (1+b)}{1-(1-z)^{b-1}}. \tag{S3.26}$$

Note that

$$\frac{\lambda'(z)}{\lambda(z)} = \frac{1}{z} \frac{1}{(1-z)\left(1-(1-z)^{b-1}\right)} \left(z(1-b) - \left((1-z)^b - (1-z)\right)\right).$$

I claim that if $b \neq 1$, $\lambda'(z)/\lambda(z) < 0$. Note that $z(1-b)$ is the support line for $(1-z)^b - (1-z)$ at $z = 0$. If $b > 1$, $(b < 1)$ then $(1-z)^b - (1-z)$ is strictly convex (concave). Thus,

$$b > 1 \Rightarrow z(1-b) - ((1-z)^b - (1-z)) < 0 \text{ and } b < 1 \Rightarrow z(1-b) - ((1-z)^b - (1-z)) > 0. \tag{S3.27}$$

Now consider the remaining part of the expression for λ'/λ. Inspection shows that

$$b > 1 \Rightarrow \frac{1}{z\left((1-z)\left(1-(1-z)^{b-1}\right)\right)} > 0 \text{ and } b < 1 \Rightarrow \frac{1}{z\left((1-z)\left(1-(1-z)^{b-1}\right)\right)} < 0. \tag{S3.28}$$

Conditions (S3.27) and (S3.28) verify that $\lambda'/\lambda < 0$. Now consider λ'. $\lambda' = \lambda\,(\lambda'/\lambda)$. When $b > 1$, $\lambda < 0$, thus $\lambda' > 0$. Similarly, if $\beta < 1$, $\lambda' < 0$. Thus, when $b > 1$, λ is λ is strictly decreasing and when $b < 1$, λ is strictly increasing. Inspection of equation (S3.25) shows that λ being strictly decreasing (increasing) implies that DRatioKumar is strictly increasing (decreasing). Thus, if $\beta > 1$, DRatioKumar is strictly increasing. Hence, The MLHR implies that the monotonicity of DRatioKumar controls the monotonicity of Λ. □

Given the result, the final demonstration is simple. The definition of R_χ given by equation (S3.21), Equation (S3.24), and Result 9.1, imply that

$$b > 1 \Rightarrow R'_\chi(y) > 0 \text{ and } b < 1 \Rightarrow R'_\chi(y) < 0.$$

Thus, for $b > 1$, χ is geometrically convex and for $b < 1$, χ is geometrically concave. This result implies by the argument given at the start of this section, that if $b > 1$ the stochastically dominant up-sized distribution is geometrically dominant and if $b < 1$ the up-sized distribution is geometrically dominated.

S3.10 Lognormal

Suppose, without loss of generality, that the common log variance of the original and upscaled Lognormal equals 1. Let r_{LN} represent the reversed hazard rate of the standard Lognormal distribution and let r_{N} and h_{N} represent the reversed hazard rate and hazard rate respectively of the standard normal distribution. Then, the definitions of the Normal and Lognormal distributions imply that

$$r_{\mathrm{LN}}(x) = \frac{1}{x} r_{\mathrm{N}}(\log(x)), \quad x > 0.$$

The symmetry of the Normal distribution implies that

$$r_{\mathrm{N}}(\log(x)) = h_{\mathrm{N}}(-\log(x)), \quad x > 0.$$

Thus,

$$\frac{x\, r'_{\mathrm{LN}}(x)}{r_{\mathrm{LN}}(x)} = -\frac{h'_{\mathrm{N}}(-\log(x))}{h_{\mathrm{N}}(-\log(x))}. \tag{S3.29}$$

First note that $x \hookrightarrow -\log(x)$ is strictly decreasing. Next, note that, as shown above in the analysis of the Normal distribution, h_{N} is strictly logconcave and thus $x \hookrightarrow h'_{\mathrm{N}}(x)/h_{\mathrm{N}}(x)$ is also strictly decreasing. Finally, note that $x \hookrightarrow -x$ is strictly decreasing. Thus, the right-hand side of (S3.29) is strictly decreasing and, hence by Lemma 14, the upscaled distribution is strictly geometrically dominant.

S3.11 Fréchet

Selection equivalence is quite easy to show using many of the tests developed in the manuscript. Perhaps the most transparent demonstration follows from noting that if G is standard $s = 1$ Fréchet with shape parameter α and F is an s-upscaling of G then, the transform function, u, is given by $u(t) = t^S$, where $S = s^\alpha$. The function $t \hookrightarrow t^S$ is clearly geometrically linear. Thus, the s-upscaling is geometrically equivalent to the original distribution and hence, by Lemma 10, the distributions are selection equivalent.

S3.12 Log Logistic

Consider the reversed hazard rate for Log-logistic distribution with shape parameter $\beta > 1$ and scale parameter $\alpha = 1$. The reversed hazard rate for this distribution is given by

$$r_{\mathrm{LLogistic}(x)} = \frac{\beta}{x + x^{1+\beta}}, \quad x > 0.$$

Using Lemma 14 to test for geometric concavity we see that

$$\frac{x\, r'_{\text{LLogistic}}(t)}{r_{\text{LLogistic}}(t)} = \left(\frac{1}{1+x^{\beta}} - 1\right)\beta - 1.$$

This expression is clearly strictly decreasing and thus, $r_{\text{LLogistic}}$ is strictly geometrically concave. Hence, the upscaled distribution always strictly geometrically dominates the original distribution.

S3.13 Gompertz

In this section let G represent the distribution function of a standard unit-scale Gompertz distribution, i.e.,

$$G(x) = 1 - \exp((1-e^x)\,\eta), \quad x > 0.$$

Let r represent the reversed hazard rate of the unit-scaled Gompertz distribution,

$$r(x) = \frac{1}{\eta}\left(e^{-x}\left(\exp\left(\eta\,(e^x - 1)\right) - 1\right)\right), \quad x > 0.$$

Suppose that F is an upscaling of G, i.e, $F(x) = G(x/s)$, $s > 1$. As in the case of the Gamma we define $\theta = 1/s$. Expressed in terms of θ, $F = G(\theta x)$, $\theta \in (0,1)$, and the reversed hazard rates of F and G are given by

$$r_F(x) = \frac{f(x)}{F(x)} = \theta\, r(\theta x); \quad r_G(x) = \frac{g(x}{G(x)} = r(x), \quad x > 0.$$

S3.13.1 Gompertz distribution when $\eta \geq 1$

When the scale parameter for the Gompertz distribution equals 1, its distribution function is given by

$$G(x) = 1 - \exp\left(\eta\,(1-e^x)\right), \quad x \in (0,\infty).$$

The reversed hazard rate for this distribution can be written as

$$\begin{aligned} r_{\text{Gompertz}}(x) &= \frac{\eta\,(1-\phi(x))\,(1-\phi(\eta\,\phi(x)))}{\phi(\eta\,\phi(x))}, \\ \phi(x) &= 1 - e^x. \end{aligned}$$

The test for Geometric concavity given in Lemma 14, requires that $x \hookrightarrow xr'(x)/r(x)$ be decreasing. Next note that

$$\begin{aligned} R_{\text{Gompertz}} = \frac{x\, r'_{\text{Gomperz}}(x)}{r_{\text{Gomperz}}} &= -\Lambda(\phi(x))\,M(\phi(x)), \\ \Lambda(y) &= e^{y\eta} - (1 + y\eta) + \eta, \quad y \leq 0, \\ M(y) &= \frac{\log(1-y)}{1-e^{y\eta}}, \quad y \leq 0. \end{aligned} \tag{S3.30}$$

Because ϕ is strictly decreasing in x, equation (S3.30) shows that to prove that R_{Gompertz} is strictly decreasing and thus r_{Gomperz} is strictly geometrically concave we need only show that the map $y \hookrightarrow \Lambda(y)\,M(y)$ is strictly decreasing.

First we show that $M$is positive and is also strictly decreasing if $\eta \geq 1$. To see this, first note that because $y < 0$, M is clearly positive. Next note that both the numerator and denominator of the expression defining M vanish as $y \to 0$ and the derivative of denominator never changes signs, thus, by the MLHR, a sufficient condition for M to be (strictly) decreasing is that

$$y \hookrightarrow \frac{\partial_y \log(1-y)}{\partial_y (1-e^{y\eta})} = \frac{e^{-y\eta}}{(1-y)\,\eta} \text{ is (strictly) decreasing.}$$

Simple differentiation shows that $e^{-y\eta}/((1-y)\,\eta)$ is strictly decreasing in y if and only if $1-\eta\,(1-y) < 0, y < 0$. This condition is satisfied if and only if $\eta \geq 1$. Thus M is positive and strictly decreasing for all $y < 0$ if and only if $\eta \leq 1$.

Now consider Λ. Λ is strictly convex over $y \leq 0$ and at $y = 0$, the upper end point of its domain of definition, $(-\infty, 0]$, Λ and its derivative both vanish. Thus, Λ is strictly decreasing in y and is positive. Thus, the product $\Lambda(y)\,M(y)$ is strictly decreasing when $\eta \geq 1$. Thus, when $\eta \geq 1$, r_{Gompertz} is strictly geometrically concave. This implies that when $\eta \geq 1$, the upscaled distribution is strictly geometrically dominant.

If on the other hand, $\eta < 1$, then explicit calculations show that

$$\lim_{y \to 0} (M(y)\,L(y))' = \frac{1-\eta}{2} > 0.$$

Thus, on a sufficiently small neighborhood of 0. $y \hookrightarrow M(y)\,L(y)$ is increasing, which implies that inverse hazard rate is not geometrically convex. Thus, when $\eta < 1$, the upscaled distribution is not geometrically dominant. We now turn characterizing selection dominance in this case.

S3.13.2 Gompertz distribution when $\eta < 1$

Result 13.1. The Gompertz distribution function is log-concave and thus r_F and r_G are strictly decreasing.

Proof. The fact that the distribution function of the Gompertz distribution is strictly logconcave, implies that r is strictly decreasing and thus that r_F and r_G are strictly decreasing. The logconcavity of the Gompertz distribution is well known (Sengupta and Nanda, 1999). □

Result 13.2. The function ρ defined by

$$\rho(x) = \frac{r_F(x)}{r_G(x)}, \quad x > 0,$$

has the following properties.
(i) $\lim_{x \to 0} \rho(x) = 1$
(ii) $\lim_{x \to \infty} \rho(x) = \infty$

Proof. Explicit calculation and, in the case of (i), the use of L'Hospital's Rule. □

Result 13.3. The derivative of ρ verifies the equation

$$\rho'(x) = \rho(x)\left(\zeta(x) - (r_F(x) - r_G(x))\right), \quad x > 0, \tag{S3.31}$$

$$\text{where } \zeta(x) = \frac{f'(x)}{f(x)} - \frac{g'(x)}{g(x)}, \quad x > 0. \tag{S3.32}$$

Proof. Differentiate using the product and quotient rules. □

Result 13.4. The function ζ defined in Result 13.3 has the following properties:
(i) $\zeta(0) = -(1-\theta)(1-\eta) < 0$ and $\lim_{x\to\infty}\zeta(x) = \infty$.
(ii) ζ is strictly increasing.

Proof. Explicit calculation in the Gompertz distribution case show that

$$\zeta(x) = e^x\left(\eta - (e^x)^{-(1-\theta)}\,\eta\,\theta\right) - (1-\theta).$$

The results then follow from calculus. □

Result 13.5. Let $x^o = \inf\{x > 0 : \rho(x) \geq 1\}$, then
(i) $\rho(x^o) = 1$.
(ii) $\zeta(x^o) > 0$.

Proof. Part (i) follows from the definition of x^o, the continuity of ρ, and Result 13.2. To prove part (ii), note that by the quotient rule for differentiation:

$$\left(\frac{F}{G}\right)' = \frac{g}{G}\left(\frac{f}{g} - \frac{F}{G}\right), \tag{S3.33}$$

$$\left(\frac{f}{g}\right)' = \frac{g}{f}\,\zeta. \tag{S3.34}$$

Let x^* be the unique 0 of ζ. Result 13.4 and (S3.34), f/g is U-shaped, strictly decreasing for $x < x^*$ and strictly increasing for $x > x^*$. Evaluating f/g at 0 and applying L'Hopitals's rule to F/G at 0, shows that $\lim_{x\to 0} F(x)/G(x) = f(0)/g(0) = \theta < 1$. Thus, $F/G < 1$ and strictly decreasing whenever $x < x^*$. Hence, $x^* < x^o$, and thus by Result 13.4, $\zeta(x^o) > 0$. □

Result 13.6. Let $x^o = \inf\{x > 0 : \rho(x) \geq 1\}$, then
(i) If $x < x^o$, then $\rho(x) < 1$.
(ii) If $x > x^o$, then $\rho(x) > 1$.

Proof. Parts (i) simply follows from the definition of x^o and the continuity of ρ. The proof of part (ii) is essentially identical to the proof of Result 5.7. The only difference is that in the Gompertz case, x^o plays the role that 0 played in the analysis of the Gamma distribution. □

Result 13.7. Let $x^o = \inf\{x > 0 : \rho(x) \geq 1\}$, then if $x \geq x^o$, ρ is strictly increasing.

Proof. The proof of part (ii) is essentially identical to the proof of Result 5.7. The only difference is that in the Gompertz case, x^o plays the role that 0 played in the analysis of the Gamma distribution. □

Define the distribution function M_x over the positive real line by

$$M_x(z) = \begin{cases} \frac{\int_0^z g(w)F(w)\,dw}{\int_0^x g(w)F(w)\,dw} & \text{if } z \leq x \\ 1 & \text{if } z > x. \end{cases} \tag{S3.35}$$

Next, note that we can write $\mathscr{R}$ (see Definition S1-3) as

$$\mathscr{R}(x) = \int_0^\infty \rho(z)\,dM_x(z), \tag{S3.36}$$

where ρ is defined in Result 13.2. For $x > x^o$, let C_x represent the distribution of M_x conditioned on $z > x^o$, where x^o is defined in Result 13.6, i.e.,

$$C_x(z) = \begin{cases} 0 & \text{if } z < x^o \\ \frac{1}{1-M_x(x^o)}\,(M_x(z) - M_x(x^o)) & \text{if } x \geq x^o. \end{cases}$$

Let $\gamma(x)$ represent the conditional expectation of ρ under C_x, i.e.,

$$\gamma(x) = \mathbb{E}_x[\rho(\tilde{z})|\tilde{z} > x^o] = \int_{x^o}^{\infty} \rho(z)\, d\, C_x(z).$$

Results 13.6 and 13.7 show that if $z \geq x^o$, $\rho > 1$ and is strictly increasing. Note that the conditional measures, C_x are ordered by strict stochastic dominance, i.e., if $x' < x''$, then $C_{x'}$ is strictly stochastically dominated by $C_{x''}$. Therefore, for $x > x^o$,

$$1 < \gamma(x) \text{ and } \gamma \text{ is strictly increasing.} \tag{S3.37}$$

Next note that using equation (S3.36), we can write

$$\mathscr{R}(x) = \int_0^{\infty} \rho(z) dM_x(z) = \int_0^{x^o} \rho(z) dM_x(z) + (1 - M_x(x^o))\gamma(x).$$

Reverting to the definition of M_x given in equation (S3.35) yields

$$\mathscr{R}(x) = \frac{\int_0^{x^o} \rho(z)\, g(z)\, F(z) dz + \gamma(x) \int_{x^o}^{x} g(z)\, F(z) dz}{\int_0^{x} g(z)\, F(z) dz}.$$

Differentiating the log of this expression yields,

$$(\log \circ \mathscr{R}(x))' = \frac{\gamma'(x) \int_0^{x} g(z)\, F(z) dz + \gamma(x)\, g(x)\, F(x)}{\int_0^{x^o} \rho(z)\, g(z)\, F(z) dz + \gamma(x) \int_{x^o}^{x} g(z)\, F(z) dz} - \frac{g(x)\, F(x)}{\int_0^{x^o} g(z)\, F(z) dz + \int_{x^o}^{x} g(z)\, F(z) dz}. \tag{S3.38}$$

Next, divide the numerator and denominator of the first term on the right-hand side of (S3.38) by γ to yield

$$(\log \circ \mathscr{R}(x))' = \frac{(\gamma'(x)/\gamma(x)) \int_0^{x} g(z)\, F(z) dz + g(x)\, F(x)}{\int_0^{x^o} (\rho(z)/\gamma(x))\, g(z)\, F(z) dz + \int_{x^o}^{x} g(z)\, F(z) dz} - \frac{g(x)\, F(x)}{\int_0^{x^o} g(z)\, F(z) dz + \int_{x^o}^{x} g(z)\, F(z) dz}. \tag{S3.39}$$

By expression (S3.37), $\gamma'(x)/\gamma(x) \geq 0$. By Results 13.6.i, and expression (S3.37), $\rho(z)/\gamma(x) < 1$ for $z \leq x^o$. Thus, the numerator of the first term on the right-hand side of equation (S3.39) exceeds the numerator of the second term and the denominator of the first term is less than the denominator of the second term. Thus, the right-hand side of equation (S3.39) is positive, which implies that for $x > x^o$, $\mathscr{R}'(x) > 0$.

Hence (a) $\mathscr{R}$ is strictly increasing for $x > x^o$. (b) For $x < x^o$, $\rho(x) < 1$, and thus for $x < x^o$, $\mathscr{R} < 1$. Because of the stochastic dominance of the upscaled Gompertz, (c) $\lim_{x\to\infty}\mathscr{R}(x) > 1$. Conditions (a), (b), and (c), imply that the supremum of $\mathscr{R}$ equals its limit as $x \to \infty$. Because for the Gompertz distribution, $\bar{x} = \infty$, Theorem 4 implies that the upscaled distribution CSSD dominates the original distribution.

S4 Formal derivation of conditional distributions used in Example 1

First note that because $\tilde{X} \overset{d}{\sim} F(x) = p + (1-p)\,G(x)$, $x \geq 0$, $\tilde{Y} \overset{d}{\sim} G \overset{d}{\sim}$ Unif(0,1), and $\tilde{X}$ and $\tilde{Y}$ are independent, we can express $\tilde{X}$ as follows

$$\tilde{X} = \tilde{B} \times 0 + (1-\tilde{B}) \times \tilde{Y}', \quad \tilde{Y}' \overset{d}{\sim} F. \tag{S4.1}$$

where $\tilde{Y}' \overset{d}{\sim} \mathrm{Unif}(0,1)$, and $\tilde{B}$ is a Bernoulli random variable equal to 1 with probability p and 0 with probability $1-p$, and $\tilde{Y}'$, $\tilde{Y}$, and $\tilde{B}$ are jointly independent.

The definition of conditional probability and the facts that $\{\tilde{X} > \tilde{Y} \;\&\; \tilde{B} = 1\}$ and $\{\tilde{X} > \tilde{Y} \;\&\; \tilde{B} = 1\}$ are disjoint events whose union equals $\{\tilde{X} > \tilde{Y}\}$ imply that the selection-conditioned distribution of $\tilde{X}$ is given by

$$\mathbb{P}[\tilde{X} \leq x | \tilde{X} > \tilde{Y}] = \frac{\mathbb{P}[\tilde{X} \leq x \,\&\, \tilde{X} > \tilde{Y}]}{\mathbb{P}[\tilde{X} > \tilde{Y}]} = \frac{\mathbb{P}[\tilde{X} \leq x \,\&\, \tilde{X} > \tilde{Y} \;\&\; \tilde{B} = 1] + \mathbb{P}[\tilde{X} \leq x \,\&\, \tilde{X} > \tilde{Y} \;\&\; \tilde{B} = 0]}{\mathbb{P}[\tilde{X} > \tilde{Y}]}. \tag{S4.2}$$

First, consider the numerator of the right-hand side of equation (S4.2). Because $\tilde{X} \leq \tilde{Y}$ whenever $\tilde{B} = 1$, and because $\tilde{X} = \tilde{Y}'$ whenever $\tilde{B} = 0$,

$$\mathbb{P}[\tilde{X} \leq x \,\&\, \tilde{X} > \tilde{Y} \;\&\; \tilde{B} = 1] + \mathbb{P}[\tilde{X} \leq x \,\&\, \tilde{X} > \tilde{Y} \;\&\; \tilde{B} = 0] = \mathbb{P}[\tilde{Y}' \leq x \,\&\, \tilde{Y}' > \tilde{Y} \;\&\; \tilde{B} = 0]. \tag{S4.3}$$

Next, note that $\tilde{B}$ is jointly independent of $\tilde{Y}$ and $\tilde{Y}'$ and $\mathbb{P}[\tilde{B} = 0] = 1 - p$. Thus,

$$\mathbb{P}[\tilde{Y}' \leq x \,\&\, \tilde{Y}' > \tilde{Y} \;\&\; \tilde{B} = 0] = (1-p)\,\mathbb{P}[\tilde{Y}' \leq x \,\&\, \tilde{Y}' > \tilde{Y}]. \tag{S4.4}$$

Next, note that, by the definition of the maximum and the fact that the common distribution function of $\tilde{Y}$ and $\tilde{Y}'$ is continuous

$$\mathbb{P}[\max[\tilde{Y}', \tilde{Y}] \leq x] = \mathbb{P}[\tilde{Y}' \leq x \,\&\, \tilde{Y}' > \tilde{Y}] + \mathbb{P}[\tilde{Y} \leq x \,\&\, \tilde{Y} > \tilde{Y}']. \tag{S4.5}$$

Because $\tilde{Y}$ and $\tilde{Y}'$ are identically distributed,

$$\mathbb{P}[\tilde{Y}' \leq x \,\&\, \tilde{Y}' > \tilde{Y}] = \mathbb{P}[\tilde{Y} \leq x \,\&\, \tilde{Y} > \tilde{Y}']. \tag{S4.6}$$

Thus, expressions (S4.5) and (S4.6) imply that

$$\mathbb{P}[\tilde{Y}' \leq x \,\&\, \tilde{Y}' > \tilde{Y}] = \tfrac{1}{2}\,\mathbb{P}[\max[\tilde{Y}', \tilde{Y}] \leq x]. \tag{S4.7}$$

Expressions (S4.3), (S4.4), and (S4.7) imply that

$$\mathbb{P}[\tilde{X} \leq x \,\&\, \tilde{X} > \tilde{Y}] = \tfrac{1}{2}\,(1-p)\,\mathbb{P}[\max[\tilde{Y}', \tilde{Y}] \leq x]. \tag{S4.8}$$

Now, consider the denominator of the right-hand side of equation (S4.2):

$$\mathbb{P}[\tilde{X} > \tilde{Y}] = \mathbb{P}[\tilde{X} > \tilde{Y} \;\&\; \tilde{B} = 1] + \mathbb{P}[\tilde{X} > \tilde{Y} \;\&\; \tilde{B} = 0] = \mathbb{P}[0 > \tilde{Y} \;\&\; \tilde{B} = 1] + \mathbb{P}[\tilde{Y}' > \tilde{Y} \;\&\; \tilde{B} = 0] = \tfrac{1}{2}\,(1-p). \tag{S4.9}$$

Expressions (S4.2), (S4.8), and (S4.9), thus imply that the selection conditioned distribution of $\tilde{X}$ is given by

$$F^c(x) = \mathbb{P}[\tilde{X} \leq x | \tilde{X} > \tilde{Y}] = \mathbb{P}[\max[\tilde{Y}', \tilde{Y}] \leq x]. \tag{S4.10}$$

Now, consider the selection-conditioned distribution of $\tilde{Y}$:

$$\mathbb{P}[\tilde{Y} \leq x | \tilde{Y} > \tilde{X}] = \frac{\mathbb{P}[\tilde{Y} \leq x \,\&\, \tilde{Y} > \tilde{X}]}{\mathbb{P}[\tilde{Y} > \tilde{X}]} = \\ \frac{\mathbb{P}[\tilde{Y} \leq x \,\&\, \tilde{Y} > \tilde{X} \,\&\, \tilde{B} = 1] + \mathbb{P}[\tilde{Y} \leq x \,\&\, \tilde{Y} > \tilde{X} \,\&\, \tilde{B} = 0]}{\mathbb{P}[\tilde{Y} > \tilde{X}]}. \tag{S4.11}$$

Note that $\tilde{X} = 0$ whenever $\tilde{B} = 1$ and that $\tilde{Y} > 0$, with probability 1. These facts and the and independence of $\tilde{B}$ imply that

$$\mathbb{P}[\tilde{Y} \leq x \,\&\, \tilde{Y} > \tilde{X} \,\&\, \tilde{B} = 1] = \mathbb{P}[\tilde{Y} \leq x \,\&\, \tilde{Y} > 0 \,\&\, \tilde{B} = 1] \\ = \mathbb{P}[\tilde{Y} \leq x \,\&\, \tilde{B} = 1] = p\,\mathbb{P}[\tilde{Y} \leq x]. \tag{S4.12}$$

$\tilde{X} = \tilde{Y}'$ whenever $\tilde{B} = 0$ and the independence of $\tilde{B}$ imply that

$$\mathbb{P}[\tilde{Y} \leq x \,\&\, \tilde{Y} > \tilde{X} \,\&\, \tilde{B} = 0] = \mathbb{P}[\tilde{Y} \leq x \,\&\, \tilde{Y} > \tilde{Y}' \,\&\, \tilde{B} = 0] = (1-p)\,\mathbb{P}[\tilde{Y} \leq x \,\&\, \tilde{Y} > \tilde{Y}']. \tag{S4.13}$$

Expression (S4.7) and the fact that $\tilde{Y}$ and $\tilde{Y}'$ are identically distributed, imply that

$$\mathbb{P}[\tilde{Y} \leq x \,\&\, \tilde{Y} > \tilde{Y}'] = \tfrac{1}{2}\,\mathbb{P}[\max[\tilde{Y}', \tilde{Y}] \leq x]. \tag{S4.14}$$

Equations (S4.11), (S4.12), (S4.13), and (S4.14), thus imply that

$$\mathbb{P}[\tilde{Y} \leq x \,\&\, \tilde{Y} > \tilde{X}] = p\,\mathbb{P}[\tilde{Y} \leq x] + \tfrac{1}{2}(1-p)\,\mathbb{P}[\max[\tilde{Y}', \tilde{Y}] \leq x]. \tag{S4.15}$$

Now consider the denominator of (S4.11):

$$\mathbb{P}[\tilde{Y} > \tilde{X}] = \mathbb{P}[\tilde{Y} > \tilde{X} \,\&\, \tilde{B} = 1] + \mathbb{P}[\tilde{Y} > \tilde{X} \,\&\, \tilde{B} = 0] = \\ \mathbb{P}[\tilde{Y} > 0 \,\&\, \tilde{B} = 1] + \mathbb{P}[\tilde{Y}' > \tilde{Y} \,\&\, \tilde{B} = 0] = p + \tfrac{1}{2}(1-p). \tag{S4.16}$$

Equations (S4.11), (S4.15), and (S4.16), thus imply that the selection conditioned distribution of $\tilde{Y}$ is given by

$$\mathbb{P}[\tilde{Y} \leq x | \tilde{Y} > \tilde{X}] = \frac{p\,\mathbb{P}[\tilde{Y} \leq x] + (1-p)\,\frac{1}{2}\,\mathbb{P}[\max[\tilde{Y}', \tilde{Y}] \leq x]}{p + \frac{1}{2}(1-p)}. \tag{S4.17}$$

The properties of the uniform distribution and its order statistics imply that

$$\mathbb{P}[\tilde{Y} \leq x] = x \quad \text{and} \quad \mathbb{P}[\max[\tilde{Y}', \tilde{Y}] \leq x] = x^2, \quad x \in [0,1]. \tag{S4.18}$$

Expressions (S4.10), (S4.17), and (S4.18) yield the expressions for the selection conditioned distributions of $\tilde{X}$ and $\tilde{Y}$ in Example 1 of the manuscript.

References

BARICZ, Á. (2010): "Geometrically concave univariate distributions," *Journal of Mathematical Analysis and Applications*, 363(1), 182–196.

CHAN, W., F. PROSCHAN, AND J. SETHURAMAN (1990): "Convex-ordering among functions, with applications to reliability and mathematical statistics," *Lecture Notes-Monograph Series*, 16, pp. 121–134.

HARDY, G. H., J. E. LITTLEWOOD, AND G. POLYA (1952): *Inequalities*. Cambridge university press.

KUCZMA, M. (2000): *An Introduction to the Theory of Functional Equations and Inequalities*. Birkhäuser Verlag.

NICULESCU, C. P., AND L.-E. PERSSON (2004): *Convex Functions and Their Applications: A Contemporary Approach*. Springer-Verlag.

PINELIS, I. (2002): "L'Hospital type rules for monotonicity, with applications," *J. Ineq. Pure & Appl. Math*, 3(1), Art. 5, 5 pp.

SAMPFORD, M. R. (1953): "Some inequalities on Mill's ratio and related functions," *The Annals of Mathematical Statistics*, 24(1), pp. 130–132.

SENGUPTA, D., AND A. K. NANDA (1999): "Log-concave and concave distributions in reliability," *Naval Research Logistics (NRL)*, 46(4), 419–433.