Stochastic Dominance Under Independent Noise*

Luciano Pomatto† Philipp Strack‡ Omer Tamuz§

September 3, 2018

1 Introduction

The notions of first and second-order stochastic dominance play a central role in the economics of information, the analysis of choice under uncertainty, and other fields. The knowledge that two random variables of interest are comparable in terms of stochastic dominance is often an important ingredient in the analysis of economic models and in obtaining unambiguous comparative statics predictions.

The purpose of this paper is to study stochastic orders when the distributions under consideration are not taken in isolation, but, as in many cases of interest, are subject to additional, statistically independent noise. The main conclusion of our analysis is that independent noise facilitates the ranking of random variables in terms of stochastic dominance.

Consider two random variables, X and Y. We ask the following question: under what conditions is it possible to find a random variable Z, independent from X and Y, so that $X + Z$ first-order stochastically dominates $Y + Z$?

Our main result is that such a (zero-mean) random variable Z exists whenever X has higher mean than Y. We further show that if X and Y have equal mean, but the first has lower variance, then Z can be chosen so that $X + Z$ dominates $Y + Z$ in terms of second-order stochastic dominance. The conditions of having higher mean or lower variance are necessary for the results to hold, and no other assumptions are imposed on X and Y.

We present applications of our findings to choice over lotteries in the presence of background risk, to the characterization of mean-variance preferences of Markowitz (1952) and Tobin (1958), and to mechanism design with risk averse agents.

*We thank Kim Border, Simone Cerreia-Vioglio, Ed Green and Massimo Marinacci for their comments. All errors and omissions are our own.

†Caltech.

‡UC Berkeley.

§Caltech. Omer Tamuz was supported by a grant from the Simons Foundation (#419427).

2 Definitions and Main Results

A random variable X *first-order stochastically dominates* Y, denoted $X \geq_1 Y$, if it satisfies $\mathbb{E}[\phi(X)] \geq \mathbb{E}[\phi(Y)]$ for every increasing function ϕ for which the two expectations are well defined. We write $X >_1 Y$ if, in addition, X and Y do not have the same distribution.[1] A random variable X *second-order* stochastically dominates Y, denoted $X \geq_2 Y$, if $\mathbb{E}[\phi(X)] \geq \mathbb{E}[\phi(Y)]$ for every *concave* increasing function ϕ for which the two expectations are well defined. As before, we write $X >_2 Y$ if X and Y have distinct distributions.

The following standard result will be useful. Given any two random variables X and Y, if $X \geq_1 Y$ then $X + Z \geq_1 Y + Z$ for any Z that is independent from the two.[2] The corresponding conclusion holds for second-order stochastic dominance. We can now state our first main result:

Theorem 1. *Let X and Y be random variables with finite expectation. If $\mathbb{E}[X] > \mathbb{E}[Y]$ then there exists a random variable Z that is independent from X and Y and such that*

$$X + Z >_1 Y + Z. \tag{1}$$

A few remarks are in order. As we show in the proof, the converse to the result is also true: if there exists a variable Z for which (1) holds, then two distributions must satisfy $\mathbb{E}[X] > \mathbb{E}[Y]$. When Z has finite expectation, this follows from the fact that any two variables that are ranked by $>_1$ are also ranked by their expectation.

It is important to point out that the conclusion of Theorem 1 is true not with respect to a single random variable Z but for a general class of distributions. The relation $X + W >_1 Y + W$ holds for any $W = Z + Z_1 + \ldots + Z_n$ that is obtained from Z by the addition of extra terms (Z_i) that are independent from X, Y and Z.

Underlying the result is the following intuition. As is well known, first-order stochastic dominance between $X + Z$ and $Y + Z$ is equivalent to the requirement that the cumulative distribution function (or cdf) F_{Y+Z} of the random variable $Y + Z$ is greater, pointwise, than the cdf F_{X+Z} of $X + Z$. The assumption that X has higher expectation than Y implies that the cdfs of the two random variables must satisfy

$$\mathbb{E}[X] - \mathbb{E}[Y] = \int_{-\infty}^{\infty} (F_Y(t) - F_X(t))\mathrm{d}t > 0 \tag{2}$$

So, on average, the function F_Y must lie above F_X. Now consider adding an independent random variable Z, distributed according to a probability density f_Z. Then, given a point

[1] Equivalently, $X >_1 Y$ if $\mathbb{E}[\phi(X)] > \mathbb{E}[\phi(Y)]$ for all strictly increasing functions ϕ for which the two expectations are well defined.

[2] If $X \geq_1 Y$ then for every increasing ϕ and $z \in \mathbb{R}$, $\mathbb{E}[\phi(X+z)] \geq \mathbb{E}[\phi(Y+z)]$ by applying the definition of FOSD to the function $x \mapsto \phi(x+z)$. Hence, $\mathbb{E}[\phi(X+Z)] = \int \mathbb{E}[\phi(X+z)]\mathrm{d}F_Z(z) \geq \int \mathbb{E}[\phi(Y+z)]\mathrm{d}F_Z(z) = \mathbb{E}[\phi(Y+Z)]$ and thus $X + Z \geq_1 Y + Z$.

$s \in \mathbb{R}$, the difference between the resulting cdfs can be expressed as

$$F_{Y+Z}(s) - F_{X+Z}(s) = \int_{-\infty}^{\infty} (F_Y(t) - F_X(t)) f_Z(s-t) \mathrm{d}t \tag{3}$$

If f_Z is sufficiently diffuse (for instance, by taking Z to be uniformly distributed around s over a sufficiently large support) then it follows from the strict inequality (2) that the difference (3) is positive at s. The crucial difficulty is to show that there exists a well-defined density such that the difference $F_{Y+Z} - F_{X+Z}$ is positive for every $s \in \mathbb{R}$ simultaneously. The existence of such a distribution F_Z is non-trivial. For example (as we argue in Section 5) Normal distributions can not in general be used to establish this result.

The proof of Theorem 1 builds on mathematical techniques introduced, in a different context, by Ruzsa and Székely (1988). In Section 4, we provide an explicit and elementary construction of the noise term Z. We show, in particular, that when X and Y have bounded support, Z can be taken to be a (random) sum of uniformly distributed random variables. While Z has infinite support, it is nevertheless well behaved. For instance, if X and Y have all moments then so will Z. More generally, as we show in the proof of Theorem 1 if X and Y have n moments then Z will have (at least) $(n-1)$ moments. Finally, if X and Y have finite variance then Z can be taken, without loss of generality, to have mean zero.

Our second main result parallels Theorem 1 and establishes an analogous conclusion for second-order stochastic dominance:

Theorem 2. *Let X and Y be random variables with finite variance. If $\mathbb{E}[X] = \mathbb{E}[Y]$ and $\mathrm{Var}[X] < \mathrm{Var}[Y]$ then there exists a random variable Z that is independent from X and Y and such that*

$$X + Z >_2 Y + Z.$$

3 Applications

3.1 Choices and Background Risk

Consider a decision maker choosing among monetary lotteries. In many situations, such a choice will be made in the presence of additional, independent, sources of risk. Prime examples include uninsurable labor risk and risk related to the value of human capital. These additional sources of risks are often not completely observable by the analyst. A large literature (e.g., Gollier and Pratt (1996), among many others) studies how the presence of such background risk affects risk aversion.

The next result, a corollary of Theorem 1, shows that large background risk can lead to risk neutral behavior: Given a finite set of gambles, there is a source of background risk such that any agent whose preferences are monotone with respect to first-order stochastic dominance will rank as more preferable gambles with higher expectation.

Corollary 1. *Let $\{X_1, \ldots, X_k\}$ be a finite collection of random variables ordered by their mean, such that $\mathbb{E}[X_i] > \mathbb{E}[X_j]$ if $i > j$. Then there exists a random variable Z independent from $\{X_1, \ldots, X_k\}$ such that $X_i + Z >_1 X_j + Z$ for all $i > j$.*

Proof. For each i there exists a random variable Z_i independent from X_i and X_{i+1} and such that $X_{i+1} + Z_i >_1 X_i + Z_i$. Without loss of generality, we can assume that each Z_i is independent from $\{X_1, \ldots, X_k\}$ and furthermore that $\{Z_1, \ldots, Z_k\}$ are independent.

Let $Z = Z_1 + \ldots + Z_k$. Then $X_{i+1} + Z_i >_1 X_i + Z_i$ implies $X_{i+1} + Z >_1 X_i + Z$. Because $>_1$ is transitive, we then obtain that $X_i + Z >_1 X_j + Z$ whenever $i > j$. □

As discussed in the previous section, the gamble Z can be assumed to have mean zero as long as each X_i has finite variance.

Corollary 1 is reminiscent of the classical Arrow-Pratt approximation. As is well known, any expected utility maximizer, when choosing out of a finite menu of monetary gambles that are sufficiently small, will behave approximately as a risk neutral agent, and select as optimal the gamble with the highest expected value.[3] Corollary 1 establishes a similar conclusion for the case where the gambles under consideration are coupled with a suitable background risk Z. It implies that a decision maker who has preferences that are consistent with first-order stochastic dominance—a class much larger than expected utility[4]—will behave like a risk-neutral expected utility maximizer when facing some (potentially large) background risks. While the Arrow-Pratt approximation applies to gambles that are small in absolute terms we look at gambles that are small relative to the background risk the decision maker is facing.

Corollary 1 extends to the case where some gambles have equal mean. Given a collection $\{X_1, \ldots, X_k\}$ where any two gambles have either distinct mean or distinct variance, there is a random variable Z independent from $\{X_1, \ldots, X_k\}$ such that $X_i + Z >_1 X_j + Z$ if $\mathbb{E}[X_i] > \mathbb{E}[X_j]$, and $X_i + Z >_2 X_j + Z$ if $\mathbb{E}[X_i] = \mathbb{E}[X_j]$ but $\mathrm{Var}[X_i] < \mathrm{Var}[X_j]$. This follows by extending the proof of Corollary 1 by combining Theorems 1 and 2. So, in the presence of the background risk Z, the decision maker will first discard gambles with lower expected payoff, and then break any remaining ties by selecting the lottery with minimal variance.

[3] More precisely, consider a set of gambles $\{kX_1, \ldots, kX_n\}$ where $k \geq 0$ measures the size of the risk. Under expected utility and a differentiable utility function, the certainty equivalent of each kX_i is given by $k\mathbb{E}[X_i]$ plus a term vanishing at rate k^2.

[4] Commonly used examples in this large class are cumulative prospect theory preferences (Tversky and Kahneman, 1992), rank dependent utility (Quiggin, 1991) and cautious expected utility (Cerreia-Vioglio et al., 2015).

3.2 Mean-Variance Preferences

In this section we use Theorem 2 to prove a simple, and to the best of our knowledge novel, axiomatization of the classic mean-variance preferences of Markowitz (1952) and Tobin (1958).

We denote by $\mathcal{P}_\infty$ the set of random variables X that have all moments. That is, $X \in \mathcal{P}_\infty$ if $\mathbb{E}[|X|^n] < \infty$ for all $n \in \mathbb{N}$. Note that we show in our proofs of Theorems 1 and 2 that when $X, Y \in \mathcal{P}_\infty$ then Z can be chosen from $\mathcal{P}_\infty$. We consider a decision maker whose preferences over monetary lotteries are described by a certainty equivalent functional $C\colon \mathcal{P}_\infty \to \mathbb{R}$ that associates to each lottery X the sure amount of money $C(X)$ that makes her indifferent between X and $C(X)$.

With slight abuse of notation, we denote by x the constant random variable that takes value $x \in \mathbb{R}$. For the next result, a random variable X is said to be a *mean preserving spread* of Y if the two have the same expectation and Y second-order stochastically dominates X.[5]

Proposition 1. *A functional $C\colon \mathcal{P}_\infty \to \mathbb{R}$ satisfies:*

1. (Certainty) $C(x) = x$ *for all* $x \in \mathbb{R}$*;*
2. (Monotonicity) *If X is a mean preserving spread of Y then $C(X) \le C(Y)$; and*
3. (Additivity) *If X, Y are independent then $C(X+Y) = C(X) + C(Y)$;*

if and only if there exists $k \ge 0$ such that

$$C(X) = \mathbb{E}[X] - k\mathrm{Var}[X].$$

Proposition 1 characterizes mean-variance preferences by three simple properties. The Certainty axiom is necessary for $C(X)$ to be interpreted as a certainty equivalent. Monotonicity requires C to rank as more desirable gambles that are less dispersed around their mean. The Additivity axiom says that the certainty equivalent is additive for independent gambles.

A key step in the proof of Proposition 1 is to show that when restricted to mean 0 random variables, a functional C that satisfies properties (1)-(3) is a decreasing function of the variance. This is an immediate implication of Theorem 2.

Proof of Proposition 1. It is immediate to verify that properties (1)-(3) are satisfied by the representation. We now prove the converse implication. We first restrict the attention to $X, Y \in \mathcal{P}_\infty$ with $\mathbb{E}[X] = \mathbb{E}[Y] = 0$, and show that $C(X) \ge C(Y)$ if and only if $\mathrm{Var}[X] \le \mathrm{Var}[Y]$.

[5] Equivalently, X is smaller than Y in the convex order.

If $\mathrm{Var}[X] < \mathrm{Var}[Y]$, by Theorem 2 there exists an independent random variable Z such that $X + Z >_2 Y + Z$. By monotonicity and additivity, this implies $C(X + Z) = C(X)+C(Z) > C(Y)+C(Z) = C(Y+Z) \Rightarrow C(X) \geq C(Y)$. Consider now the case where $\mathrm{Var}[X] = \mathrm{Var}[Y] = \sigma^2$. We show that $C(X) = C(Y)$. Assume, as way of contradiction, that $C(X) - C(Y) = \varepsilon > 0$. Let $(X_1, X_2, \dots)$ and $(Y_1, Y_2, \dots)$ be i.i.d. sequences that are independent from X and Y, and with the properties that each X_i is distributed as X, and each Y_i is distributed as Y. Because $X_i \geq_2 X$ and $X \geq_2 X_i$, then $C(X_i) = C(X)$. Similarly, $C(Y_i) = C(Y)$. By additivity,

$$C(X_1 + \cdots + X_n) - C(Y_1 + \cdots + Y_n + X + Y) = \varepsilon n - C(X) - C(Y),$$

which is positive for n large enough. However, $\sigma^2 n = \mathrm{Var}[X_1 + \cdots + X_n] < \mathrm{Var}[Y_1 + \cdots + Y_n + X + Y] = \sigma^2(n + 2)$, and so by the first part of the proof we have that $C(X_1 + \ldots + X_n) - C(Y_1 + \ldots + Y_n + X + Y) \leq 0$ and we reached a contradiction. Hence $C(X) = C(Y)$.

Thus, restricted to zero mean $X \in \mathcal{P}_\infty$, C satisfies $C(X) = f(\mathrm{Var}[X])$ for some monotone decreasing function f. Furthermore, f is additive (i.e., $f(x+y) = f(x)+f(y)$) since C is additive. As is well known, every monotone additive $f\colon \mathbb{R}_+ \to \mathbb{R}$ is linear. To see this, note that if $m, n \in \mathbb{N}$, then, letting $q = m/n$, we have $mf(1) = f(m) = f(nq) = nf(q)$. Thus, $f(q) = qf(1)$ for every rational q. Hence, for irrational x, if we let $\lim_n q_j = x$ for a monotone increasing sequence q_j, then $f(x) \leq f(q_j) \leq \lim_j f(q_j) = xf(1)$. Considering a monotone decreasing sequence converging to x yields $f(x) \geq xf(1)$. Hence $f(x) = xf(1)$ for every x. Hence $C(X) = f(1)\mathrm{Var}[X]$ for every mean 0 random variable X. To conclude the proof, notice that for any X (with possibly non-zero mean) additivity implies $C(X - \mathbb{E}[X]) = C(X) - \mathbb{E}[X]$. Therefore, letting $k = f(1)$, we obtain $C(X) = \mathbb{E}[X] - k\mathrm{Var}[X]$. □

We end this section with an additional, similar result characterizing the expectation as the unique functional on $\mathcal{P}_\infty$ that is monotone with respect to first order stochastatic dominance, and additive for independent random variable.

Proposition 2. *A functional $C\colon \mathcal{P}_\infty \to \mathbb{R}$ satisfies:*

1. (Monotonicity) *If $X >_1 Y$ then $C(X) \geq C(Y)$; and*

2. (Additivity) *If X, Y are independent then $C(X + Y) = C(X) + C(Y)$;*

if and only if there exists $k \geq 0$ such that

$$C(X) = k\mathbb{E}[X].$$

Proof of Proposition 2. This proof closely follows that of Proposition 1. As in that proof, it is immediate to verify that properties (1)-(2) are satisfied by the representation.

Denote (as above) by x the random variable that take the value $x \in \mathbb{R}$ with probability 1, and define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = C(x)$. By additivity and monotonicity f is monotone increasing and additive and so, as in the proof of Proposition 1, $f(x) = kx$ for some $k \geq 0$.

We show that if $\mathbb{E}[X] > \mathbb{E}[Y]$ then $C(X) \geq C(Y)$. Indeed, in this case, by Theorem 1 there is an independent $Z \in \mathcal{P}_\infty$ such that $X + Z >_1 Y + Z$, and so by monotonicity $C(X + Z) \geq C(Y + Z)$. By additivity it follows that $C(X) \geq C(Y)$. Hence for any $\varepsilon > 0$ it holds that

$$C(X) \leq C(\mathbb{E}[X] + \varepsilon) = f(\mathbb{E}[X] + \varepsilon)$$

and

$$C(X) \geq C(\mathbb{E}[X] - \varepsilon) = f(\mathbb{E}[X] - \varepsilon).$$

Thus $C(X) = f(\mathbb{E}[X]) = k\mathbb{E}[X]$. □

3.3 Implementation with Risk Aversion

In this section we show how Theorem 1 can be used to construct mechanisms that are robust to uncertainty about the agents' risk attitudes.

Consider a mechanism design problem with n agents, where each agent's type is $\theta_i \in \Theta_i$, and $\theta = (\theta_1, \dots, \theta_n)$ is drawn from some joint distribution. We assume that the set of types $\Theta = \prod_i \Theta_i$ is finite. The designer chooses an allocation $x \in X$ and a transfer $t_i \in \mathbb{R}$ for each agent i. By the revelation principle we can without loss of generality restrict to mechanisms where each agent reports their type and it is optimal for the agents to be truthful. A direct mechanism (x, t) is a tuple consisting of an allocation function $x : \prod_{i=1}^n \Theta_i \to X$ and a transfer function $t : \prod_{i=1}^n \Theta_i \to \Delta(\mathbb{R}^n)$. We restrict attention to mechanisms where each random transfer $t_i(\theta)$ has all moments. Most of the literature on mechanism design focuses on the quasi-linear case, where agent i's utility is given by

$$v_i(x, \theta) - t_i\,.$$

Here, $v_i(x, \theta)$ denotes the monetary equivalent of the utility agent i derives from the physical allocation x when the type profile equals $\theta = (\theta_1, \dots, \theta_n)$. Note, that as v depends on the complete type profile we allow for interdependent values.

A restriction which is imposed when assuming quasi-linear preferences is that the agents are risk neutral over money. To more generally model agents who might not be risk neutral, we assume that preferences are of the form

$$u_i\left(v_i(x, \theta) - t_i\right)\,, \tag{4}$$

where u_i is agent i's (nondecreasing) utility function over money. These preferences are commonly considered in the literature on auctions with risk aversion (for example case 1. in Maskin and Riley, 1984) or in the literature on optimal income taxation (see for example Diamond, 1998).

Given such preferences a mechanism (x, t) is *Bayes incentive compatible* if for every type θ_i and every agent i it is optimal to report her type truthfully to the mechanism, given that all other players do the same

$$\theta_i \in \arg\max_{\hat{\theta}_i} \mathbb{E}\Big[u_i\left(v_i(x(\hat{\theta}_i, \theta_{-i}), \theta) - t_i(\hat{\theta}_i, \theta_{-i})\right) \mid \theta_i\Big] \tag{u-BIC}$$

A mechanism is Bayes incentive compatible (BIC) *under quasi-linear preferences* if it satisfies (u-BIC) when all agents' utilities over money are linear, i.e. $u_i(x) = x$ for all agents i. We say that a mechanism is strictly Bayes incentive compatible if the maximum in (u-BIC) is unique.

Note, that Bayes incentive compatibility depends on the utility functions $u = (u_1, \ldots, u_n)$ and thus the risk attitudes of the agents. A mechanism which is IC for moderately risk averse agents might not be IC for risk neutral or very risk averse agents. This poses a problem in situations where the designer might not have precise knowledge of the risk attitudes. While most of the literature focuses on the case where the utility functions of the agents' are known to the designer we are interested in finding mechanisms (x, t) which implement a given physical allocation x robustly for any vector of (potentially different) utilities $u = (u_1, \ldots, u_n)$. To explicitly model this problem we define the stronger notion of *robust incentive compatibility* which requires that it is optimal for the agents to be truthful independently of their risk attitudes:

Definition 1. [Robust incentive compatibility] A direct mechanism (x, t) is robustly incentive compatible if (u-BIC) is satisfied for all non-decreasing utility functions $u_1, u_2, \ldots, u_n$.

It is natural to ask which physical allocation rules x can be implemented robustly.[6] Our main result in this section is that (essentially) any allocation that can be implemented when the agents have quasi-linear utilities (and the designer knows this) can also be implemented when the agents have arbitrary utilities u that are unknown to the designer. Thus, perhaps surprisingly, knowledge of the agents risk attitudes is not necessary to implement a given allocation.

Proposition 3. *Suppose the direct mechanism (x, t) is strictly Bayes incentive compatible with quasi-linear preferences. Then, there exist another* ***robust incentive compatible*** *direct mechanism (x, τ) that implements the same physical allocation x and raises the same expected revenue from each agent given each vector of reported types $\hat{\theta}$, i.e., $\mathbb{E}[\tau_i(\hat{\theta})] = \mathbb{E}[t_i(\hat{\theta})]$ for every i.*

Proof. The proof of Proposition 3 is a direct application of Corollary 1. Fix agent i's type to be θ_i. For each $\hat{\theta}_i \in \Theta_i$ denote by $\mu_{\hat{\theta}_i}$ the distribution of $v_i(x(\hat{\theta}_i, \theta_{-i}), \theta) - t_i(\hat{\theta}_i, \theta_{-i})$

[6] A first observation is that every mechanism which is ex-post or dominant strategy IC is also robustly IC.

(i.e., the monetary utility of i when reporting $\hat{\theta}_i$) conditioned on the true type θ_i. Let $X_{\hat{\theta}_i}$ be distributed according to $\mu_{\hat{\theta}_i}$. Strict Bayes incentive compatibility says that

$$\mathbb{E}[X_{\theta_i}] > \mathbb{E}[X_{\hat{\theta}_i}]$$

whenever $\hat{\theta}_i \neq \theta_i$. By Corollary 1 there thus exists an independent random variable Z_{θ_i} such that $X_{\theta_i} + Z_{\theta_i} >_1 X_{\hat{\theta}_i} + Z_{\theta_i}$. Furthermore, since the random variables $(X_{\hat{\theta}_i})$ have all moments, we can assume that $Z_{\theta_i'}$ does too, and has zero mean. By repeating this construction for each player i and each true type θ_i, we obtain a collection (Z_{θ_i}) of random variables which we can assume to be independent. Let $Z_i = \sum_{\theta_i \in \Theta_i} Z_{\theta_i}$. Then it still holds that $X_{\theta_i} + Z_i >_1 X_{\hat{\theta}_i} + Z_i$ for every every θ_i, and so the mechanism (x, τ) with $\tau_i(\hat{\theta}) = t_i(\hat{\theta}) + Z_i$ is robust incentive compatible. In addition, $\mathbb{E}[\tau_i(\hat{\theta})] = \mathbb{E}[t_i(\hat{\theta})]$ since Z_i has zero mean. □

Intuitively, Theorem 1 says that if a random outcome is preferred on average then it is preferred in the sense of first order stochastic dominance, under some background risk. Hence, to make a Bayes IC mechanism robustly IC it suffices to add a carefully chosen risk to the transfer that is independent of the agents' reports.

A few comments on Proposition 3 are in order. First, the mechanism (x, τ) is a randomized mechanism in the sense that the transfers payed by the agents are random – even conditional on all agents' types. Second, as (x, t) and (x, τ) implement the same physical allocation x, the set of strictly implementable physical allocations is the same under Bayes IC and robust IC. This is in contrast to dominant strategy IC for which the set of implementable physical allocations is strictly smaller.[7]

A second immediate observation is that robust incentive compatibility not only implies that it is optimal for an agent with expected utility preferences to report truthfully, but for any preference that respects first-order stochastic dominance over lotteries. Examples include prospect theory and rank-dependent utility.

While Proposition 3 holds only for strictly implementable allocation rules this is often a minor restriction (with finitely many types) as in many standard setups for any Bayes implementable allocation rule there is a strictly implementable allocation rule that raises almost as much revenue. Consider, for example, the case where $X \subseteq \mathbb{R}^n$, $v_i(x, \theta) = x_i \cdot H(\theta_i)$ and $H_i : \Theta_i \to \mathbb{R}_+$ is increasing in θ_i. This case covers most of the models analyzed in the literature, for example many public good problems as well as auction and allocation problems. For this case, we show in Lemma 2 in the Appendix that for

[7]For example, in a single good allocation problem it is well known that a physical allocation rule x is strictly implementable in a Bayes IC mechanism if and only if it is monotone on average, i.e. $z \mapsto \mathbb{E}[x_i(\theta) \mid \theta_i = z]$ is increasing and dominant strategy IC if and only if it is realization by realization monotone, i.e. $\theta_i \mapsto x_i(\theta_i, \theta_{-i})$ is increasing for all i, θ_{-i}. Note, that as Gershkov et al. (2013) show, the set of implementable interim utilities under some conditions is the same under dominant strategy and Bayesian incentive compatibility even if the set of implementable allocation rules differs.

every Bayes IC mechanism (x, t) and every $\epsilon > 0$ there exists another strictly Bayes IC mechanism (x, τ) that implements the same physical allocation x and raises at most ϵ less in expected revenue $\mathbb{E}[\tau_i(\theta)] \geq \mathbb{E}[t_i(\theta)] - \epsilon$.

In our analysis we have abstracted away from individual rationality. We note that a mechanism that is individually rational when agents are risk neutral is not necessarily individually rational when u_i is made concave (the simplest example being a common value auction). Furthermore, the addition of background noise may violate individual rationality for risk averse agents, and would thus require the agents to be compensated to ensure their participation. This, however, while affecting the transfers would not change the set of implementable physical allocation rules.

4 Uniformity

In this section we address the following two questions. First, is the noise term obtained in Theorem 1 robust to changes in the distribution of X and Y? Moreover, can it be given a simple description? By restricting the attention to random variables with bounded support, we provide positive answers to both questions.

We consider pairs of random variables X and Y such that:

(a) Their support is included in a bounded interval $[-M, M]$; and

(b) Their difference in mean is bounded below by $\mathbb{E}[X] - \mathbb{E}[Y] \geq \varepsilon M$, where $\varepsilon > 0$.

Given M and ε, we construct a variable Z that satisfies $X + Z >_1 Y + Z$ for *any* pair X and Y for which (a) and (b) hold. In addition, we show that Z can be taken to be a combination of uniformly distributed random variables.

The random variable Z is defined by three parameters: M and ε, as described above, as well as a parameter $a > 0$, which for the next result we are going to take to be sufficiently large. Let $U_1, U_2, \ldots$ be i.i.d. random variables that are uniformly distributed on the union of intervals $[-a-M, -a+M] \cup [a-M, a+M]$. Let N be an independent geometric random variable with parameter $1/2$, so that $\mathbb{P}[N = n] = 2^{-1-n}$ for $n \in \{0, 1, 2, \ldots\}$. Finally, let R_1 and R_2 be variables independent from N and $U_1, U_2, \ldots$ and uniformly distributed on $[-a, a]$. We define

$$Z = R_1 + R_2 + (U_1 + U_2 + \cdots U_N). \tag{5}$$

So, the random variable Z is obtained as a sum of mean-zero, independent noise terms that are uniformly distributed.

Theorem 3. *Fix $M, \varepsilon > 0$ and $a \geq 16M\varepsilon^{-1} + 8M$. The random variable Z, as defined in (5), is such that for every X, Y supported in $[-M, M]$ with $\mathbb{E}[X] - \mathbb{E}[Y] \geq \varepsilon M$ it holds that*

$$X + Z >_1 Y + Z.$$

Notice that for smaller ε, as the difference in expectation between X and Y becomes negligible, the support of each term U_i becomes increasingly large. The random variable Z is reasonably "well-behaved": for example, it has all moments, and exponentially vanishing tails. Using Wald's Lemma, its variance can be shown to be

$$\mathrm{Var}[Z] = \frac{2}{3}(M/\varepsilon)^2 \left(1024 + 1024\varepsilon + 257\varepsilon^2\right),$$

so that its standard deviation is of order $M\varepsilon^{-1}$, and never more than $30M\varepsilon^{-1}$.

To put this into perspective, consider an agent who must make a choice between two lotteries that pay between -\$10 and \$10, and whose expected value differs by at least \$1. Theorem 3 implies that there exists a zero mean independent background noise which—for *any* utility function—makes the lottery with the higher expectation preferable. Moreover, this noise need not be incredibly large: a standard deviation of \$3000 suffices.

5 Discussion and Limitations

Our main result establishes a bridge between two orderings—having strictly greater mean and first-order stochastic dominance—that differ substantially in their strength and implications. We conclude by discussing some of the limitations of our result.

We first observe that the noise term Z in Theorem 1 cannot be bounded, unless additional assumptions are imposed on X and Y. Indeed, in order for $X + Z >_1 Y + Z$ to be satisfied, the maximum of the support of $X + Z$ must lie weakly above that of $Y + Z$. In particular, the same relation must hold between X and Y.

Finally, one may wonder whether a construction simpler than the one we offered in Theorem 3 would suffice. For example, can Z be taken to be Gaussian? It so happens that it is impossible for a Gaussian Z to satisfy Theorem 1; this holds even if we restrict the random variables in question to be finitely supported. Indeed, for any Z that satisfies Theorem 3 it must hold that $\mathbb{E}[\exp(tZ)] = \infty$ for some $t \in \mathbb{R}$ large enough.[8] This holds because if the maximum of the support of X is strictly less than that of Y (which of course does not preclude that $\mathbb{E}[X] > \mathbb{E}[Y]$) then $\mathbb{E}[\exp(tX)] < \mathbb{E}[\exp(tY)]$ for t large enough. Hence for any independent Z for which $\mathbb{E}[\exp(tZ)]$ is finite it holds that

$$\mathbb{E}\left[\mathrm{e}^{t(X+Z)}\right] = \mathbb{E}\left[\mathrm{e}^{tX}\right] \cdot \mathbb{E}\left[\mathrm{e}^{tZ}\right] < \mathbb{E}\left[\mathrm{e}^{tY}\right] \cdot \mathbb{E}\left[\mathrm{e}^{tZ}\right] = \mathbb{E}\left[\mathrm{e}^{t(Y+Z)}\right],$$

and so it cannot be that $X + Z >_1 Y + Z$.

[8] Note that many standard distributions satisfy this: for example any geometric distribution, exponential distribution or gamma distribution. All of these distributions have all moments and exponentially vanishing tails, as does our Z.

References

S. Cerreia-Vioglio, D. Dillenberger, and P. Ortoleva. Cautious expected utility and the certainty effect. *Econometrica*, 83(2):693–728, 2015.

P. A. Diamond. Optimal income taxation: An example with a u-shaped pattern of optimal marginal tax rates. *The American Economic Review*, 88(1):83–95, 1998.

D. H. Fremlin. *Measure theory*, volume 2. Torres Fremlin, 2002.

A. Gershkov, J. K. Goeree, A. Kushnir, B. Moldovanu, and X. Shi. On the equivalence of bayesian and dominant strategy implementation. *Econometrica*, 81(1):197–220, 2013.

C. Gollier and J. W. Pratt. Risk vulnerability and the tempering effect of background risk. *Econometrica*, pages 1109–1123, 1996.

H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

E. Maskin and J. Riley. Optimal auctions with risk averse buyers. *Econometrica*, 52(6): 1473–1518, 1984.

L. Mattner. What are cumulants? *Documenta Mathematica*, 4:601–622, 1999.

L. Mattner. Cumulants are universal homomorphisms into Hausdorff groups. *Probability theory and related fields*, 130(2):151–166, 2004.

J. Quiggin. Comparative statics for rank-dependent expected utility theory. *Journal of Risk and Uncertainty*, 4(4):339–350, 1991.

S. T. Rachev, L. Klebanov, S. V. Stoyanov, and F. Fabozzi. *The methods of distances in the theory of probability and statistics*. Springer, 2013.

I. Ruzsa and G. J. Székely. *Algebraic probability theory*. John Wiley & Sons Inc, 1988.

J. Tobin. Liquidity preference as behavior towards risk. *The Review of Economic Studies*, 25(2):65–86, 1958.

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

A Proof of Theorem 1

Denote by $\mathcal{P}_n$ the collection of all Borel probability measures on $\mathbb{R}$ that have finite nth moment:

$$\mathcal{P}_n = \left\{\nu \,:\, \int_{\mathbb{R}} |x|^n \,\mathrm{d}\nu(x) < \infty\right\}.$$

We denote $\mathcal{P}_\infty = \bigcap_n \mathcal{P}_n$. Likewise, denote by $\mathcal{M}_n$ the collection of all bounded Borel signed measures on $\mathbb{R}$ that have finite nth moment. That is, $\mu \in \mathcal{M}_n$ if $\int_{\mathbb{R}} |x|^n \,\mathrm{d}\mu(x) < \infty$. Recall that a signed measure μ is bounded if its absolute value $|\mu|$ is a finite measure. As usual, given $\mu, \nu \in \mathcal{M}_n$ we write $\mu \geq \nu$ if $\mu(A) \geq \nu(A)$ for every Borel $A \subseteq \mathbb{R}$. We equip $\mathcal{M}_n$ with the total-variation norm

$$||\mu|| = \int_{\mathbb{R}} 1 \,\mathrm{d}|\mu|.$$

We denote the convolution of $\mu_1, \mu_2 \in \mathcal{M}_n$ (and in its subset $\mathcal{P}_n$) by $\mu_1 * \mu_2$, and by $\mu^{(k)}$ the k-fold convolution of μ with itself for all $k \geq 1$. To simplify notation we define $\mu^{(0)}$ to be δ, the Dirac measure at zero, so that $\mu^{(k)} * \mu^{(m)} = \mu^{(k+m)}$ for all $k, m \geq 0$. Note that $\mathcal{M}_n$, equipped with the norm defined above, is a Banach space which is closed under convolutions. In fact, $(\mathcal{M}_n, *, +)$ is a Banach algebra, so that $\mu * (\nu_1 + \nu_2) = \mu * \nu_1 + \mu * \nu_2$.

The following lemma is due to Ruzsa and Székely (1988, pp. 126–127). It states that a signed measure that assigns total mass 1 to $\mathbb{R}$ can be "smoothed" into a probability measure by convolving it with an appropriately chosen probability measure. We provide the proof for the reader's convenience; an essentially identical proof also appears in Mattner (1999, p. 616), as well as in Mattner (2004, p. 159).

Lemma 1 (Ruzsa and Székely, Mattner)**.** *Let $n \in \{1, 2, \ldots, \infty\}$. For every $\mu \in \mathcal{M}_n$ with $\mu(\mathbb{R}) = 1$ there is a $\nu \in \mathcal{P}_n$ such that $\mu * \nu \in \mathcal{P}_n$.*

Proof. Let ρ_a be the uniform probability distribution on $[-a, a]$. For some $0 < c < 1$ and some $\pi \in \mathcal{P}_n$ let

$$\tau = (1 - c) \sum_{k=0}^{\infty} c^k \pi^{(k)},$$

where $\pi^{(0)}$ is δ, the Dirac measure at zero. Since $c < 1$ the series converges and so τ is a probability measure. Let

$$\nu = \rho_a^{(2)} * \tau.$$

We show that $\mu * \nu \in \mathcal{P}_n$ for an appropriate choice of a, c and π. To see that $\mu * \nu$ is a probability measure note first that $[\mu * \nu](\mathbb{R}) = 1$,[9] and so it suffices to show that $\mu * \nu$ is

[9]This follows immediately from the fact that $\int_{-\infty}^{\infty} \mathrm{d}(\mu * \nu)(x) = \int_{-\infty}^{\infty} \mathrm{d}\mu(x) \times \int_{-\infty}^{\infty} \mathrm{d}\nu(x)$.

positive. To see this, we write

$$\begin{aligned}\mu * \nu &= \rho_a^{(2)} * \mu * \tau \\ &= \rho_a^{(2)} * (\mu - \delta + \delta - c\pi + c\pi) * \tau \\ &= \rho_a^{(2)} * (\mu - \delta + c\pi) * \tau + \rho_a^{(2)} * (\delta - c\pi) * \tau.\end{aligned}$$

Now, by the definition of τ,

$$c\pi * \tau = (1-c) \sum_{k=0}^{\infty} c^{k+1} \pi^{(k+1)} = \tau - (1-c)\delta.$$

Hence $(\delta - c\pi) * \tau = (1-c)\delta$, and so

$$\mu * \nu = \rho_a^{(2)} * (\mu - \delta + c\pi) * \tau + (1-c)\rho_a^{(2)}.$$

Now, note that $\rho_a^{(2)} \geq \frac{1}{2}\rho_a$. Hence

$$\begin{aligned}\rho_a^{(2)} * (\mu - \delta + c\pi) &\geq \rho_a^{(2)} * (\mu - \delta) + \frac{c}{2}\rho_a * \pi && (6)\\ &= \frac{c}{2}\rho_a * \left[\frac{2}{c}\rho_a * (\mu - \delta) + \pi\right]. && (7)\end{aligned}$$

Thus, if we choose a, c and π so that $\frac{2}{c}\rho_a * (\mu - \delta) + \pi$ is a positive measure it will follow that $\mu * \nu$ is also positive. This is indeed the case if we set

$$\frac{c}{2} = ||\rho_a * (\mu - \delta)||.$$

and

$$\pi = \frac{2}{c}|\rho_a * (\mu - \delta)| = \frac{1}{||\rho_a * (\mu - \delta)||}|\rho_a * (\mu - \delta)|,$$

and if $0 < c < 1$. If $c = 0$ for some a then $\rho_a * \mu = \rho_a$ and we can take $\nu = \rho_a$ to conclude the proof of the theorem. In addition, $c = 2||(\mu - \delta) * \rho_a||$ tends to 0 as a tends to infinity (see Mattner, 2004, 2.5 (d)), so we can choose a large enough so that $c < 1$.

We have shown that $\mu * \nu$ is a sum of positive measures, hence positive, and hence a probability measure. It remains to be shown that ν has finite nth moment, and hence is in $\mathcal{P}_n$. To this end, note first that

$$\pi = \frac{2}{c}|\rho_a * (\mu - \delta)| \leq \frac{2}{c}(\rho_a * (|\mu| + \delta)) \in \mathcal{M}_n,$$

and so $\pi \in \mathcal{P}_n$. Since the nth moment of $\pi^{(k)}$ is at most k^n times the nth moment of π, it follows that $\tau \in \mathcal{P}_n$, and so $\nu = \rho_a^{(2)} * \tau \in \mathcal{P}_n$. □

We now proceed with the proof of Theorem 1. Let X and Y be two random variables with $c = \mathbb{E}[X] - \mathbb{E}[Y] > 0$. Define the signed measure σ as

$$\sigma(A) = \frac{1}{c}\int_A F_Y(t) - F_X(t)\,\mathrm{d}t$$

for every Borel $A \subseteq \mathbb{R}$. A standard application of Tonelli's Theorem implies that

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-] = \int_0^\infty 1 - F_X(t)\mathrm{d}t - \int_{-\infty}^0 F_X(t)\mathrm{d}t$$

and so

$$\mathbb{E}[X] - \mathbb{E}[Y] = -\int_\infty^0 (F_X(t) - F_Y(t))\mathrm{d}t + \int_0^\infty (F_Y(t) - F_X(t))\mathrm{d}t = \int_\mathbb{R} F_Y(t) - F_X(t)\,\mathrm{d}t.$$

Hence $\sigma(\mathbb{R}) = 1$. Furthermore, σ is a bounded measure, since

$$\begin{aligned}
c|\sigma|(\mathbb{R}) &= \int_{-\infty}^\infty |F_Y(t) - F_X(t)|\,\mathrm{d}t \\
&= \int_{-\infty}^0 |F_Y(t) - F_X(t)|\,\mathrm{d}t + \int_0^\infty |(1 - F_Y(t)) - (1 - F_X(t))|\,\mathrm{d}t \\
&\leq \int_{-\infty}^0 [F_Y(t) + F_X(t)]\,\mathrm{d}t + \int_0^\infty [(1 - F_Y(t)) + (1 - F_X(t))]\,\mathrm{d}t \\
&= \mathbb{E}[|X|] + \mathbb{E}[|Y|].
\end{aligned}$$

Also, if $X, Y \in \mathcal{P}_n$ then $\sigma \in \mathcal{M}_{n-1}$, since, using integration by parts (i.e., Tonelli's Theorem),

$$c\int_{-\infty}^\infty n|x|^{n-1}\,\mathrm{d}\sigma(x) = -\int_{-\infty}^\infty |x|^n\,\mathrm{d}(F_Y - F_X)(x) = \mathbb{E}[|X|^n] - \mathbb{E}[|Y|^n] < \infty.$$

Hence Lemma 1 implies that there exists a probability measure $\eta \in \mathcal{P}_{n-1}$ such that $\sigma * \eta \in \mathcal{P}_{n-1}$. Let Z be a random variable independent from X and Y with distribution η. The measure σ is by definition absolutely continuous with density $c(F_Y - F_X)$. Therefore, $\sigma * \eta$ is absolutely continuous as well, and its density s satisfies, almost everywhere,[10]

$$\begin{aligned}
s(x) &= c\int_\mathbb{R} F_Y(x-t) - F_X(x-t)\,\mathrm{d}\eta(t) \\
&= c\mathbb{P}[Y \leq x - Z] - c\mathbb{P}[X \leq x - Z] \\
&= c(F_{Y+Z}(x) - F_{X+Z}(x)).
\end{aligned}$$

Because $\sigma * \eta$ is a probability measure and $c > 0$, then $F_{Y+Z}(x) - F_{X+Z}(x) \geq 0$ for almost every x. Since F_{Y+Z} and F_{X+Z} are right-continuous, this implies $F_{Y+Z} \geq F_{X+Z}$. Furthermore, this inequality is strict somewhere, since the integral of $c(F_{Y+Z} - F_{X+Z})$ is equal to one. Therefore, $X + Z >_1 Y + Z$. This concludes the proof of Theorem 1; we have furthermore demonstrated that when $X, Y \in \mathcal{P}_n$ then Z can be taken to be in $\mathcal{P}_{n-1}$, for any $n \in \{1, 2, \ldots, \infty\}$.

We end this section by showing that the converse of Theorem 1 is also true: if $\mathbb{E}[X] \leq \mathbb{E}[Y]$ then there does not exist a random variable Z such that $X + Z >_1 Y + Z$. To see this, note that in this case the measure

$$\sigma'(A) = \int_A F_Y(t) - F_X(t)\,\mathrm{d}t$$

[10] See Fremlin (2002, 257xe).

satisfies $\sigma'(\mathbb{R}) \leq 0$, and so for any probability measure η it holds that $[\sigma' * \eta](\mathbb{R}) \leq 0$. It thus follows from the calculation above that $F_{Y+Z}(x) - F_{X+Z}(x) < 0$ for some x (except in the trivial case in which X and Y have the same distribution). Thus it is impossible that $X + Z >_1 Y + Z$.

B Proof of Theorem 2

Let X and Y be random variables in $\mathcal{P}_n$ with $\mathbb{E}[X] = \mathbb{E}[Y]$ and $c = \mathrm{Var}[Y] - \mathrm{Var}[X] > 0$. Let F_X and F_Y be the cumulative distribution functions of X and Y. Define the signed measure σ as

$$\sigma(A) = \frac{1}{c}\int_A \int_{-\infty}^{t} F_X(u) - F_Y(u)\,\mathrm{d}u\,\mathrm{d}t.$$

By using the assumptions that $\mathbb{E}[X] = \mathbb{E}[Y]$, $\mathbb{E}[X^2], \mathbb{E}[Y^2] < \infty$ and applying integration by parts, it follows that $\int_{\mathbb{R}} |\int_{-\infty}^{t} F_X(u) - F_Y(u)\,\mathrm{d}u\,|\mathrm{d}t < \infty$ (Rachev et al., 2013, Lemma 15.2.1). Hence σ is bounded.

A calculation similar to the one used in Theorem 1 shows that $\sigma(\mathbb{R}) = 1$ and that $\sigma \in \mathcal{M}_{n-1}$. As in the proof of Theorem 1, we can invoke Lemma 1 to prove the existence of a probability measure $\eta \in \mathcal{P}_{n-1}$ such that $\sigma * \eta \in \mathcal{P}_{n-1}$. Let Z be a random variable independent from X and Y with distribution η. Then $s(x)$, the probability density function of $\sigma * \eta$, is

$$\begin{aligned} s(x) &= c\int_{\mathbb{R}} \int_{-\infty}^{x-t} F_X(u) - F_Y(u)\,\mathrm{d}u\,\mathrm{d}\eta(t) \\ &= c\int_{\mathbb{R}} \int_{-\infty}^{x} F_X(u-t) - F_Y(u-t)\,\mathrm{d}u\,\mathrm{d}\eta(t) \\ &= c\int_{-\infty}^{x} \int_{\mathbb{R}} F_X(u-t) - F_Y(u-t)\,\mathrm{d}\eta(t)\,\mathrm{d}u \\ &= c\int_{-\infty}^{x} F_{X+Z}(u) - F_{Y+Z}(t)\,\mathrm{d}u. \end{aligned}$$

Since $\sigma * \eta$ is a probability measure then $s(x)$ is non-negative for almost every $x \in \mathbb{R}$. Since F_{Y+Z} and F_{X+Z} are right-continuous, this implies $s \geq 0$. Furthermore, this inequality is strict somewhere, since the integral of s is equal to one. Therefore, $X + Z >_2 Y + Z$, as $\int_{-\infty}^{x} F_{X+Z}(u) - F_{Y+Z}(t)\,\mathrm{d}u \geq 0$ for all x is a well known condition for second-order stochastic dominance.

C A Lemma on Strict Bayesian Implementation

Lemma 2. *Suppose that $x \in \mathbb{R}^n$ and $v_i(x, \theta) = x_i H_i(\theta_i)$ where $H_i : \Theta_i \to \mathbb{R}_+$ is strictly increasing in θ_i and that types are independently drawn. Then for every Bayes IC mechanism (x, t) and every $\epsilon > 0$ there exists another strictly Bayes IC mechanism (x, τ) that* *implements the same physical allocation x and raises at most ϵ less in expected revenue* $\mathbb{E}[\tau_i(\theta)] \geq \mathbb{E}[t_i(\theta)] - \epsilon$.

Proof. Agent i's interim utility when she is of type θ_i, but reports to be of type θ_i' can be written as

$$\mathbb{E}[v_i(x(\theta_i', \theta_{-i}), (\theta_i, \theta_{-i})) - t_i(\theta_i', \theta_{-i})] = Q_i(\theta_i')H_i(\theta_i) - T_i(\theta_i')$$

where $Q_i(\theta_i') = \mathbb{E}[x(\theta_i', \theta_{-i})]$, $T_i(\theta_i') = \mathbb{E}[t_i(\theta_i', \theta_{-i})]$. As the mechanism is Bayes incentive compatible, i.e. satisfies (u-BIC) when u equals the identity, we have that for all $\theta_i, \theta_i' \in \Theta_i$

$$Q_i(\theta_i)H_i(\theta_i) - T_i(\theta_i) \quad - \quad Q_i(\theta_i')H_i(\theta_i) + T_i(\theta_i') \geq 0 \tag{8}$$

$$Q_i(\theta_i')H_i(\theta_i') - T_i(\theta_i') \quad - \quad Q_i(\theta_i)H_i(\theta_i') + T_i(\theta_i) \geq 0\,. \tag{9}$$

Adding the two equations yield

$$\Big(Q_i(\theta_i) - Q_i(\theta_i')\Big)\Big(H_i(\theta_i) - H_i(\theta_i')\Big) \geq 0\,.$$

As H_i increases in θ_i it follows from the above equation that Q_i is non-decreasing. Furthermore, (8) and (9) imply that

$$(Q_i(\theta_i) - Q_i(\theta_i'))H_i(\theta_i') \leq T_i(\theta_i) - T_i(\theta_i') \leq (Q_i(\theta_i) - Q_i(\theta_i'))H_i(\theta_i)\,. \tag{10}$$

We will next argue that the monotonicity of Q_i in addition with (10) imposed for adjacent types $\theta_i' < \theta_i$ are also sufficient for incentive compatibility.[11] Consider an agent who is of type θ_i, but who deviates to report that she is of type θ_i', which is not necessarily adjacent. Her loss in interim expected utility from this deviation is given by

$$\begin{aligned} & Q_i(\theta_i)H_i(\theta_i) - T_i(\theta_i) - Q_i(\theta_i')H_i(\theta_i) + T_i(\theta_i') \\ = \quad & \Big(Q_i(\theta_i) - Q_i(\theta_i')\Big)H_i(\theta_i) - \Big(T_i(\theta_i) - T_i(\theta_i')\Big)\,. \end{aligned}$$

Suppose now that $\theta_i' < \theta_i$ and let $(\theta^k)_{k \in \{0,\dots,m\}}$ be a sequence of adjacent types such that $\theta_i' = \theta^0 < \theta^1 < \ldots < \theta^m = \theta_i$. We have that the gain from deviating is given by

$$\begin{aligned} & \Big(Q_i(\theta_i) - Q_i(\theta_i')\Big)H_i(\theta_i) - \Big(T_i(\theta_i) - T_i(\theta_i')\Big) \\ = \quad & \sum_{k=1}^{m} \Big(Q_i(\theta^k) - Q_i(\theta^{k-1})\Big)H_i(\theta^m) - \Big(T_i(\theta^k) - T_i(\theta^{k-1})\Big) \end{aligned}$$

By the monotonicity of Q_i and H_i we can bound the loss from below by

$$\begin{aligned} & Q_i(\theta_i)H_i(\theta_i) - T_i(\theta_i) - Q_i(\theta_i')H_i(\theta_i) + T_i(\theta_i') \\ \geq \quad & \sum_{k=1}^{m} \Big(Q_i(\theta^k) - Q_i(\theta^{k-1})\Big)H_i(\theta^k) - \Big(T_i(\theta^k) - T_i(\theta^{k-1})\Big)\,. \end{aligned} \tag{11}$$

[11] We say that two types θ_i, θ_i' are adjacent if there exists no other type θ_i'' such that $\theta_i < \theta_i'' < \theta_i'$.

As we imposed (10) for adjacent types we know that each component of the sum is non-negative. Hence, the sum is non-negative and no downward deviation is profitable. The case of upward deviations, i.e. $\theta_i' > \theta_i$, is completely analogous. We thus have that the mechanism (x, t) is Bayes incentive compatible if and only if Q_i is non-decreasing for every agent i and in addition for every two adjacent types $\theta_i' < \theta_i$ the following equation holds

$$(Q_i(\theta_i) - Q_i(\theta_i'))H_i(\theta_i') \leq T_i(\theta_i) - T_i(\theta_i') \leq (Q_i(\theta_i) - Q_i(\theta_i'))H_i(\theta_i) . \tag{12}$$

Thus, an upper bound on the transfer is given by the transfer that solves the right inequality with equality

$$T_i(\theta_i) - T_i(\theta_i') = (Q_i(\theta_i) - Q_i(\theta_i'))H_i(\theta_i) .$$

Note, that for a sufficiently small ϵ the transfer τ_i that for adjacent types $\theta_i > \theta_i'$ solves

$$\tau_i(\theta_i) - \tau_i(\theta_i') = (Q_i(\theta_i) - Q_i(\theta_i'))H_i(\theta_i) - \frac{\epsilon}{|\Theta_i|} .$$

and assigns the same transfer $\tau_i(\underline{\theta}_i) = T_i(\underline{\theta}_i)$ to the lowest type $\underline{\theta}_i$ as the original transfer T_i solves (12) with strict inequality. By (11) this implies that the mechanism is strictly incentive compatible. The mechanism (x, τ) satisfies $\tau_i(\theta_i) \geq T_i(\theta_i) - \epsilon$ and thus raises at most ϵ less revenue from each agent. Furthermore, each agent strictly prefers to be truthful over deviating and reporting an adjacent type. □

D Proof of Theorem 3

To prove Theorem 3 we first prove a version of Lemma 1 that gives a stronger results for a smaller class of measures. Given $M, L > 0$, denote by $\mathcal{M}_M^L$ the set of bounded signed measures μ that are supported in $[-M, M]$, and for which $|\mu|([a, b]) \leq L(a - b)$ for all $a > b$ in $[-M, M]$.

Proposition 4. *For every $M, L > 0$ there is a measure $\nu \in \mathcal{P}_\infty$ so that $\mu * \nu \in \mathcal{P}_\infty$ for every $\mu \in \mathcal{M}_M^L$ with $\mu(\mathbb{R}) = 1$.*

Given $\mu \in \mathcal{M}_\infty$ and $a > 0$, define

$$\mu_a = (\mu - \delta) * \rho_a,$$

where, as in the proof of Lemma 1, δ is the point mass at 0, and ρ_a is the uniform distribution on $[-a, a]$. Let $r_a = \frac{1}{2a}\mathbb{1}_{\{[-a,a]\}}$ be the density of ρ_a. It follows that μ_a has a density $m_a(x) = \int_{\mathbb{R}} r_a(x - t)\mathrm{d}(\mu - \delta)(t)$ given by

$$m_a(x) = \frac{1}{2a}\mu([x - a, x + a]) - r_a(x) \tag{13}$$

and which satisfies the following properties:

Lemma 3. *For any $\mu \in \mathcal{M}_M^L$ it holds that*

1. $|m_a(x)|$ *is at most* $ML/a + 1/(2a)$.
2. $m_a(x)$ *vanishes outside of the union of the intervals* $[-a - M, -a + M]$ *and* $[a - M, a + M]$.

Proof.
1. Since $\mu \in \mathcal{M}_M^L$, its norm $||\mu||$ is at most $2ML$, and so $|\mu([x-a, x+a])| \leq 2ML$. And since $|r_a| \leq 1/(2a)$, it follows that $|m_a| \leq ML/a + 1/(2a)$.
2. For x such that $-a + M < x < a - M$ we have that $\mu([x - a, x + a]) = 1$, since $[x-a, x+a]$ includes $[-M, M]$ and thus all of the support of μ. Since $r_a(x) = 1/(2a)$ in this range it follows that $m_a(x) = 0$. For $x < -a - M$ or $x > a + M$ we have that $r_a(x) = 0$, and that likewise $\mu([x - a, x + a]) = 0$, since now $[x + a, x - a]$ does not intersect $[-M, M]$. Hence also in this range $m_a(x) = 0$.

□

Proof of Proposition 4. The proof will follow the proof of Lemma 1 and will refer to some of the arguments given there. As in that proof, let ρ_a be the uniform distribution on $[-a, a]$. Fix some $0 < c < 1$ and $\pi \in \mathcal{P}_\infty$, and let

$$\tau = (1 - c) \sum_{k=0}^{\infty} c^k \pi^{(k)}$$

and

$$\nu = \rho_a^{(2)} * \tau.$$

Choose any $\mu \in \mathcal{M}_M^L$. Then as in the proof of Lemma 1 we have that

$$\mu * \nu = \rho_a^{(2)} * (\mu - \delta + c\pi) * \tau + (1 - c)\rho_a^{(2)}$$

and that

$$\begin{aligned} \rho_a^{(2)} * (\mu - \delta + c\pi) &\geq \frac{c}{2}\rho_a * \left[\frac{2}{c}\rho_a * (\mu - \delta) + \pi\right] \\ &= \frac{c}{2}\rho_a * \left[\frac{2}{c}\mu_a + \pi\right]. \end{aligned}$$

We now show how to choose a, c and π so that $\frac{2}{c}\mu_a + \pi$ is a positive measure, independent of our choice of $\mu \in \mathcal{M}_M^L$. By part 1 of Lemma 3 we know that the density of $\frac{2}{c}|\mu_a|$ is bounded from above by $2ML/(ac) + 1/(ac)$, and by part 2 of the same lemma we know that it vanishes outside $[-a - M, -a + M] \cup [a - M, a + M]$. Therefore, if we choose π to be the uniform distribution on this union of intervals then $\frac{2}{c}\mu_a + \pi$ will be positive if the density of π on its support—which equals $1/(4M)$—is larger than our bound on the density of $\frac{2}{c}|\mu_a|$. That is, we would like a and c to satisfy

$$\frac{1}{4M} \geq \frac{2ML}{ac} + \frac{1}{ac},$$

while keeping $c < 1$. Rearranging yields

$$ac \geq 8M^2L + 4M,$$

which is satisfied if we take $c = 1/2$ and any $a \geq 16M^2L + 8M$. The proof that $\nu \in \mathcal{P}_\infty$ is identical to the one in the proof of Lemma 1. □

Given Proposition 4, the proof of Theorem 3 follows closely the proof of Theorem 1. If X and Y are supported on $[-M, M]$ and if $\mathbb{E}[X] - \mathbb{E}[Y] \geq M\varepsilon$ then the measure σ, which is given by

$$\sigma(A) = \frac{1}{\mathbb{E}[X] - \mathbb{E}[Y]} \int_A F_Y(t) - F_X(t)\,\mathrm{d}t,$$

is in $\mathcal{M}_M^{1/(\varepsilon M)}$. Proposition 4 shows we can find $\nu \in \mathcal{P}_\infty$ such that $\sigma * \nu \in \mathcal{P}_\infty$. In addition, as shown in the proof of the same proposition, given $a \geq 16M\varepsilon^{-1} + 8M$, ν can be taken to be the distribution of the random variable Z defined as (5) in the main text. Finally, the same argument used in the proof of Theorem 1 shows that $X + Z >_1 Y + Z$.