Stochastic Dominance Under Independent Noise[*]

Luciano Pomatto[†]   Philipp Strack[‡]   Omer Tamuz[§]

May 20, 2019

Abstract

Stochastic dominance is a crucial tool for the analysis of choice under risk. It is typically analyzed as a property of two gambles that are taken in isolation. We study how additional independent sources of risk (e.g. uninsurable labor risk, house price risk, etc.) can affect the ordering of gambles. We show that, perhaps surprisingly, background risk can be strong enough to render lotteries that are ranked by their expectation ranked in terms of first-order stochastic dominance. We extend our results to second order stochastic dominance, and show how they lead to a novel, and elementary, axiomatization of mean-variance preferences.

1   Introduction

A choice between risky prospects is often difficult to make. It is perhaps even harder to predict, since decision makers vary in their preferences. One exception is when the prospects are ordered in terms of stochastic dominance. In this case, a standard prediction is that the dominant option is chosen. For this reason, stochastic dominance has long been seen as a central concept in decision theory and the economics of information, and remains an area of active research (see, e.g. Müller et al., 2016; Cerreia-Vioglio et al., 2016, among others). More generally, stochastic dominance is an important tool for non-parametric comparison of distributions.

In the typical analysis of choices under risk, stochastic dominance is studied as a property of two prospects X and Y that are taken in isolation, without considering other risks the decision maker might be facing. However, a choice between two gambles or assets

often happens in the presence of unavoidable background risks, such as risks related to health, human capital, or exposure to the aggregate uncertainty in the economy.

In this paper we study how stochastic dominance is affected by background noise. Given two random variables X and Y, we are interested in understanding if an independent risk Z can make the resulting variables $X + Z$ and $Y + Z$ ordered in terms of stochastic dominance, and, if so, under what conditions. Our main result is that such a 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 show that the distribution of Z can be obtained as a convex combination $(1 - \varepsilon)G + \varepsilon H$, where G is the distribution of a Gaussian random variable with mean zero and (possibly large) variance, and H is the distribution of an additional noise term with heavy tails. The weight ε can be taken to be arbitrarily small. Hence, up to a vanishing degree of error, the background risk is indistinguishable from Gaussian/Normal noise. The role played by H is to produce a distribution displaying sufficiently thick tails, a feature that, as we show, is necessary for the background noise to lead to stochastic dominance.

Interestingly, Acemoglu et al. (2017) provide evidence of key macroeconomic variables—presumably related to individual risk—that take a form similar to the one described above: their distributions are approximately Gaussian around their mean but display heavier tails. Our results, therefore, draw a connection between the analysis of choice under risk and the study of distributions with thick tail, a subject of long and renewed interest across different fields of economics, including macroeconomics (Gabaix et al., 2006; Morris and Yildiz, 2016), finance (Gabaix et al., 2003; Kelly and Jiang, 2014), as well as other disciplines (Nair et al., 2013).

As is well-known, ranking risky prospects solely in terms of their mean and variance is a crude approach for decision-making under risk, especially if compared to expected utility theory. Nevertheless, mean and variance remain key statistics driving the decisions of practitioners and investors. Our results show that such a seemingly ad-hoc approach can be justified in the presence of suitable background risk: while first and second order stochastic dominance are much stronger orders than the comparisons of means and variances, our results suggest that the latter are, under background risk, surprisingly good proxies for the former.

To the best of our knowledge, the only other paper to study the effect of independent noise on stochastic dominance is the contemporaneous, and independent, work of Tarsney (2019), who provides interesting examples of binary gambles that become ranked in terms of stochastic dominance after the addition of suitable background risk.

The paper is organized as follows. Section 2 presents the main results. Section 3 applies

our results to provide a simple axiomatization of mean-variance preferences, distinct from the previous approaches (Maccheroni, Marinacci, and Rustichini, 2006). In Section 4 we apply our results to the study of mechanism design with risk averse agents. In Section C in the appendix we restrict attention to bounded gambles and provide a more quantitative version of our main result.

2 Definitions and Main Results

A random variable (or gamble) X *first-order stochastically dominates* a random variable 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 X and Y have distinct distributions (equivalently, if $X \geq_1 Y$ but $Y \not\geq_1 X$).[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.

We make a few observations which will be useful in what follows. Given 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] Hence, two variables that are ranked in terms of stochastic dominance remain so after the addition of independent noise. Given a random variable Z, we say that Z' is *noisier than Z* if $Z' = Z + W$ for some random variable W that is independent of Z. Hence, if $X + Z >_1 Y + Z$ for some Z that is independent of X and Y, then $X + Z' >_1 Y + Z'$ for any independent Z' that is noisier than Z. The analogous conclusions hold for second-order stochastic dominance.

We denote by $\mathcal{P}_\infty$ the set of random variables that have all finite moments. That is, a gamble X belongs to $\mathcal{P}_\infty$ if $\mathbb{E}[|X|^n]$ is finite for all $n \in \mathbb{N}$. 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*

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

In particular, $X + Z' >_1 Y + Z'$ for every Z' that is independent of X and Y and noisier than Z. Additionally, if $X, Y \in \mathcal{P}_\infty$ then Z can be taken to belong to $\mathcal{P}_\infty$.

By taking ϕ to be linear, 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

---

[1]Yet again 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 first-order stochastic dominance 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 $X + Z \geq_1 Y + Z$.

expectation, this follows from the fact that any two variables that are ranked by $>_1$ are also ranked by their expectation.

The next result, an immediate 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 gambles in* $\mathcal{P}_\infty$ *such that* $\mathbb{E}[X_i] > \mathbb{E}[X_j]$ *if* $i > j$. *Then there is an independent* $Z \in \mathcal{P}_\infty$ *such that*

$$X_i + Z >_1 X_j + Z \quad for\ all\ \ i > j.$$

Proof. For each i there is an gamble $Z_i \in \mathcal{P}_\infty$ independent of X_i and X_{i+1} such that $X_{i+1} + Z_i >_1 X_i + Z_i$. We can further assume that each Z_i is independent of $(Z_j)_{j \neq i}$ and of $(X_1, \ldots, X_k)$. Let $Z = Z_1 + \cdots + Z_k$. Then, as Z is noisier than Z_i, $X_{i+1} + Z_i >_1 X_i + Z_i$ implies $X_{i+1} + Z >_1 X_i + Z$. Because $>_1$ is transitive, $X_i + Z >_1 X_j + Z$ whenever $i > j$. $\square$

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.

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*

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

In particular, $X + Z' >_2 Y + Z'$ *for every* Z' *that is independent of* X *and* Y *and noisier than* Z. *Additionally if* $X, Y \in \mathcal{P}_\infty$ *then* Z *can be taken to belong to* $\mathcal{P}_\infty$.

---

[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).

2.1 Proof Sketch and Discussion

Underlying Theorem 1 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 gamble $Y + Z$ is greater, pointwise, than the cdf F_{X+Z} of $X + Z$.

The assumption that X has higher expectation than Y implies, using integration by parts, 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 cumulative distribution 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 $s \in \mathbb{R}$, 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 in a neighborhood of s.

So, the crucial difficulty in establishing Theorem 1 is to show the existence of a well-defined distribution such that the difference $F_{Y+Z} - F_{X+Z}$ is positive everywhere. The existence of such a distribution F_Z is not trivial. The proof of Theorem 1 provides an explicit construction, building on mathematical techniques introduced, in a different context, by Ruzsa and Székely (1988).

While the details of the construction are somewhat technical, the background risk in Theorem 1 can be approximated (in terms of total variation distance) by Gaussian noise, up to a vanishing degree of error:

Remark 1. *Let X and Y be as in Theorem 1. For every $\varepsilon \in (0, 1)$ there is a $\sigma^2 > 0$ and a distribution with cdf H such that if G is the cdf of a Gaussian distribution with zero mean and variance σ^2, and if Z is independent of X and Y and distributed according to*

$$F_Z = (1 - \varepsilon)G + \varepsilon H,$$

then Z satisfies (1).

In the proof of Theorem 1, the background risk Z is defined as follows. We construct a sequence $U_1, U_2, \ldots$ of i.i.d. random variables and an independent geometric random variable N such that $\mathbb{P}[N = n] = (1 - \varepsilon)\varepsilon^n$ for $n \in \{0, 1, 2, \ldots\}$. Letting W be an independent Gaussian random variable, we define

$$Z = W + (U_1 + U_2 + \cdots + U_N). \tag{4}$$

So, the random variable Z is obtained as a sum of mean-zero Gaussian noise and geometric sum of independent noise terms. With probability $1 - \varepsilon$ the variable N takes value 0 and therefore Z reduces to standard Gaussian noise. With probability $(1 - \varepsilon)\varepsilon^n$, n additional terms $U_1 + \ldots + U_n$ contribute to the background risk.

2.2 Orders of Magnitude

An immediate question, at this point, is to understand the orders of magnitude involved. In particular, how large does the background risk need to be? We consider, as a representative example, a decision maker who is confronted with a small gamble X with positive expectation, and the choice of whether to take X or to receive 0 for sure. We show that a background risk Z that is not implausibly large in many contexts (e.g. risk of unemployment or risk stemming from variation in house prices) suffices to make X dominant.

In particular, suppose X pays \$12 and $-\$10$ with equal probabilities. Then, there exists a background risk Z, of the form described in (4), that satisfies $Z + X >_1 Z$, and has the following properties: its distribution is a mixture that gives weight $1 - \varepsilon = 0.99$ to a Gaussian with standard deviation \$3,500. The standard deviation of Z itself is \$3,525. If X pays \$100 and -\$50 with equal probability, then we can set $1 - \varepsilon = 0.978$ and the standard deviation of Z is \$4,551.[5]

3 Mean-Variance Preferences

In this section we apply Theorem 2 to provide a simple axiomatization of the classic mean-variance preferences of Markowitz (1952) and Tobin (1958).[6] 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 gamble 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.[7]

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 $C(X) \leq C(Y)$; and*

---

[5]Further examples can be found in Table 1 in the Appendix. Because the proof is Theorem 1 is constructive, these quantities can be easily computed numerically.

[6]A previous axiomatization appears in Epstein (1985), for DARA preferences. Related results are due to Maccheroni et al. (2006, 2009).

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

3. (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

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

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.

Suppose $\mathbb{E}[X] = \mathbb{E}[Y]$ and $\mathrm{Var}[X] < \mathrm{Var}[Y]$. Then, by Theorem 2, there exists an independent random variable Z such that $X + Z >_2 Y + Z$. By monotonicity and additivity, this implies $C(X) + C(Z) = C(X + Z) \geq C(Y + Z) = C(Y) + C(Z)$. Hence, $C(X) \geq C(Y)$.

Suppose $\mathbb{E}[X] = \mathbb{E}[Y]$ and $\mathrm{Var}[X] = \mathrm{Var}[Y] = \sigma^2$. We show that $C(X) = C(Y)$. Assume, as a way of contradiction, that $C(X) - C(Y) = \varepsilon > 0$. Let $(X_1, X_2, \ldots)$ and $(Y_1, Y_2, \ldots)$ be i.i.d. sequences where each X_i is distributed as X, and each Y_i is distributed as Y. Fix a random variable Z independent from the two sequences, and such that $\mathbb{E}[Z] = 0$ and $\mathrm{Var}[Z] > 0$. 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,

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

which is strictly positive for n large enough. However, $n\sigma^2 + \mathrm{Var}[Z] = \mathrm{Var}[X_1 + \cdots + X_n + Z] > \mathrm{Var}[Y_1 + \cdots + Y_n] = n\sigma^2$, and so, by the first part of the proof, we have that $C(X_1 + \ldots + X_n + Z) - C(Y_1 + \ldots + Y_n) \leq 0$ 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 function f. In addition, by the first part of the proof, f is nonincreasing. Furthermore, f is additive (i.e., $f(x + y) = f(x) + f(y)$) since C is additive. As is well known, every nonincreasing additive $f \colon \mathbb{R}_+ \to \mathbb{R}$ is linear.[8] To conclude the proof, notice that for any X (with possibly non-zero mean) additivity and certainty imply $C(X - \mathbb{E}[X]) = C(X) - \mathbb{E}[X]$. Therefore, letting $k = -f(1)$, we obtain $C(X) = \mathbb{E}[X] - k\mathrm{Var}[X]$. $\qquad\square$

---

[8] 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 every $x \in \mathbb{R}_+$ and positive rational q, q' such that $q \leq x \leq q'$ we have $q'f(1) = f(q') \leq f(x) \leq f(q) = qf(1)$. So, $f(x) = xf(1)$ for every x.

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

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

1. *(Certainty) $C(1) = 1$;*

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

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

if and only if $C(X) = \mathbb{E}[X]$.

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 $f\colon \mathbb{R} \to \mathbb{R}$ 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 $k \geq 0$. Certainty implies $k = 1$.

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 $C(X) = f(\mathbb{E}[X]) = \mathbb{E}[X]$. $\qquad\square$

4 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 $\theta = (\theta_1, \ldots, \theta_n)$ 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

$$u_i \left(v_i(x, \theta) - t_i \right) \,, \tag{5}$$

where $u_i : \mathbb{R} \to \mathbb{R}$ is agent i's utility function over money. We assume u_i is bounded by a polynomial; see Section 5 for a discussion of this integrability assumption. Preferences (5) are commonly considered in the literature on auctions with risk aversion (see for example case 1. in Maskin and Riley, 1984) or in the literature on optimal income taxation (see for example Diamond, 1998).

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

$$\theta_i \in \operatorname*{argmax}_{\hat{\theta}_i} \mathbb{E}\Big[u_i \left(\, v_i(x(\hat{\theta}_i, \theta_{-i}), \theta) - t_i(\hat{\theta}_i, \theta_{-i}) \right) \mid \theta_i\Big] \qquad (u\text{-BIC})$$

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. $u_i(x) = x$. A mechanism is *strictly* Bayes incentive compatible if the maximum in (u-BIC) is unique.

Bayes incentive compatibility depends on the utility functions $u = (u_1, \dots, u_n)$ and thus the risk attitudes of the agents. We are interested in finding mechanisms which implement a given physical allocation x for any vector of utilities. To explicitly model this problem we consider the stronger notion of *ordinal incentive compatibility* (d'Aspremont and Peleg, 1988).

Definition 1. A direct mechanism (x, t) is ordinal incentive compatible if (u-BIC) is satisfied for all non-decreasing utility functions $u_1, \dots, u_n$.

It is natural to ask which physical allocation rules x can be implemented ordinally.[9] Our main result in this section is that (essentially) any allocation that can be implemented when the agents have quasi-linear utilities can also be implemented when the agents have

---

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

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. To make a Bayes IC mechanism ordinally IC it suffices to add a carefully chosen risk to the transfer that is independent of the agents' reports.

Proposition 3. *Suppose the direct mechanism (x, t) is strictly Bayes incentive compatible with quasi-linear preferences. Then, there exists another **ordinally 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 $v_i(x(\hat{\theta}_i, \theta_{-i}), \theta) - t_i(\hat{\theta}_i, \theta_{-i})$ (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 ordinal incentive compatible. In addition, $\mathbb{E}[\tau_i(\hat{\theta})] = \mathbb{E}[t_i(\hat{\theta})]$ since Z_i has zero mean. $\square$

The result has the following additional implication. As (x, t) and (x, τ) implement the same physical allocation x, it follows that the set of strictly implementable physical allocations is the same under Bayes IC and ordinal IC. This is in contrast to dominant strategy IC for which the set of implementable physical allocations is strictly smaller.[10]

In our analysis we have abstracted away from individual rationality. A mechanism that is individually rational when agents are risk neutral is not necessarily individually rational when u_i is concave. The addition of background noise may violate individual rationality for risk averse agents, and thus require the agents to be compensated to ensure their participation. This, however, would not change the set of implementable physical allocation rules.

---

[10]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.

5 Discussion

Unbounded Support and Tail Risk   We first observe that the noise term Z in Theorem 1 cannot have bounded support, unless additional assumptions are imposed on X and Y. This holds even if we only consider bounded 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, if Z has bounded support and $X + Z >_1 Y + Z$, then the maximum of the support of X must be greater than the maximum of the support Y.

In addition, the background risk Z must generally display non-negligible risk at the tails. For instance, it is impossible for a Gaussian Z to satisfy Theorem 3; this holds even if we restrict the random variables in question to be finitely supported. Indeed, any Z that satisfies Theorem 3 must have thick tails in the sense that $\mathbb{E}[\exp(tZ)] = \infty$ for some $t \in \mathbb{R}$ large enough.

To see the necessity for thick tailed distributions, note that 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

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

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

Integrability   As is well known, distributions with unbounded support require the specification of a class of utility functions for which all expectations are finite, so that issues similar to the St. Petersburg paradox do not arise.

A standard solution is to restrict attention to utility functions that are bounded. See, e.g., Aumann (1977). More generally, following Russell and Seo (1978) (see also Ryan, 1974; Arrow, 1974), one may wish to consider utility functions that have polynomial tails. That is, utility functions $u : \mathbb{R} \to \mathbb{R}$ that are bounded, in absolute value, by a polynomial. This assumption is more general and allows for utility functions that are strictly increasing and strictly concave everywhere.[11] For any such u and gamble $X \in \mathcal{P}_\infty$ with all finite moments, the resulting expected utility $\mathbb{E}[u(X)]$ is well-defined and finite.

By definition, distributions with thick tails rule out CARA as a class of utility functions for which integrability is preserved. This, however, does not make thick tailed distributions pathological. Many standard distributions have thick tails: examples include geometric, exponential and gamma distributions. All of these have finite moments and exponentially vanishing tails. Distribution with thick tails have been used to describe many economic

---

[11]Given any utility function $u : \mathbb{R} \to \mathbb{R}$ and an arbitrarily large bounded interval, there exists a utility v that agrees with u on that interval and can be taken to be bounded or with polynomial tails. So, both assumptions amount to non-falsifiable integrability conditions.

variables of interest (Acemoglu et al., 2017), and are the subject of a growing literature in economics (see, e.g., Morris and Yildiz, 2016).

Calibration   Rabin's (2000) calibration theorem famously highlighted a tension between expected utility and risk aversion over small gambles. Subsequent literature clarified that the conclusions of Rabin's impossibility theorem extend to a broader class of preferences. In particular, Barberis et al. (2006) showed how independent background risk can eliminate first-order risk aversion, a feature that in the absence of background risk can, under some conditions, accommodate risk-aversion over small gambles. Our paper contributes to this literature by showing how background risk can lead to risk-neutral behavior, for any monotone preference, and for a suitable choice of background risk.

Conclusion   Our main theorem establishes a connection between two orderings—having strictly greater mean and first-order stochastic dominance—that differ substantially in their strength and implications. An interesting avenue for future research is to study more quantitative versions of our result and to what extent the stark predictions implied by our results are reflected in actual behavior.

References

D. Acemoglu, A. Ozdaglar, and A. Tahbaz-Salehi. Microeconomic origins of macroeconomic tail risks. *American Economic Review*, 107(1):54–108, 2017.

K. J. Arrow. The use of unbounded utility functions in expected-utility maximization: Response. *The Quarterly Journal of Economics*, 88(1):136–138, 1974.

R. J. Aumann. The st. petersburg paradox: A discussion of some recent comments. *Journal of Economic Theory*, 14(2):443–445, 1977.

N. Barberis, M. Huang, and R. H. Thaler. Individual preferences, monetary gambles, and stock market participation: A case for narrow framing. *American economic review*, 96 (4):1069–1090, 2006.

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

S. Cerreia-Vioglio, F. Maccheroni, and M. Marinacci. Stochastic dominance analysis without the independence axiom. *Management Science*, 63(4):1097–1109, 2016.

C. d'Aspremont and B. Peleg. Ordinal bayesian incentive compatible representations of committees. *Social Choice and Welfare*, 5(4):261–279, 1988.

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.

L. G. Epstein. Decreasing risk aversion and mean-variance analysis. *Econometrica: Journal of the Econometric Society*, pages 945–961, 1985.

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

X. Gabaix, P. Gopikrishnan, V. Plerou, and H. E. Stanley. A theory of power-law distributions in financial market fluctuations. *Nature*, 423(6937):267, 2003.

X. Gabaix, P. Gopikrishnan, V. Plerou, and H. E. Stanley. Institutional investors and stock market volatility. *The Quarterly Journal of Economics*, 121(2):461–504, 2006.

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.

B. Kelly and H. Jiang. Tail risk and asset prices. *The Review of Financial Studies*, 27 (10):2841–2871, 2014.

F. Maccheroni, M. Marinacci, and A. Rustichini. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74(6):1447–1498, 2006.

F. Maccheroni, M. Marinacci, A. Rustichini, and M. Taboga. Portfolio selection with monotone mean-variance preferences. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 19(3):487–521, 2009.

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.

S. Morris and M. Yildiz. Crises: Equilibrium shifts and large shocks, 2016. Working paper.

A. Müller, M. Scarsini, I. Tsetlin, and R. L. Winkler. Between first-and second-order stochastic dominance. *Management Science*, 63(9):2933–2947, 2016.

J. Nair, A. Wierman, and B. Zwart. The fundamentals of heavy-tails: properties, emergence, and identification. In *ACM SIGMETRICS Performance Evaluation Review*, volume 41, pages 387–388. ACM, 2013.

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

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

S. Rachev and S. Stoyanov. Advanced stochastic models, risk assessment, and portfolio optimization, the frank j. fabozzi series, 2008.

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

W. R. Russell and T. K. Seo. Admissible sets of utility functions in expected utility maximization. *Econometrica: Journal of the Econometric Society*, pages 181–184, 1978.

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

T. M. Ryan. The use of unbounded utility functions in expected-utility maximization: comment. *The Quarterly Journal of Economics*, 88(1):133–135, 1974.

M. Shaked and J. G. Shanthikumar. *Stochastic orders*. Springer Science & Business Media, 2007.

C. Tarsney. Exceeding expectations: Stochastic dominance as a general decision theory. 2019.

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 $n \in \{0, 1, \ldots, \infty\}$. 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 measure of a Gaussian random variable with mean zero and standard deviation $a > 0$.[12] 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 $[\mu * \nu](\mathbb{R}) = 1$,[13] and so it suffices to show that $\mu * \nu$

---

[12]One could take here any other distribution (e.g., the uniform probability distribution on $[-a, a]$) such that $\rho_a^{(2)} \geq \beta \rho_a$ for some $\beta > 0$.

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

is 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

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

It follows by comparing the densities of $\rho_a^{(2)}$ and ρ_a that $\rho_a^{(2)} \geq \frac{1}{\sqrt{2}}\rho_a$. Hence

$$
\rho_a^{(2)} * (\mu - \delta + c\pi) \geq \rho_a^{(2)} * (\mu - \delta) + \frac{c}{\sqrt{2}}\rho_a * \pi \tag{6}
$$

$$
= \frac{c}{\sqrt{2}}\rho_a * \left[\frac{\sqrt{2}}{c}\rho_a * (\mu - \delta) + \pi \right]. \tag{7}
$$

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

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

and

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

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, $c = \sqrt{2}\|(\mu - \delta) * \rho_a\|$ tends to 0 as a tends to infinity,[14] so we can choose a large enough so that $c < 1$, and in fact c as small as we like.

___

[14]Let δ_y be the point mass at y. Then $\delta_y * \rho_a$ is a Gaussian distribution with mean y and standard deviation a. The Kullback-Leibler divergence between the two is well-known to be $D(\delta_y * \rho_a \| \rho_a) = \frac{1}{2}\left(\frac{y}{a}\right)^2$. By Pinsker's Inequality,

$$
\|\delta_y * \rho_a - \rho_a\| \leq \sqrt{2D(\delta_y * \rho_a \| \rho_a)} = \frac{|y|}{a}.
$$

Hence $\|\delta_y * \rho_a - \rho_a\| \leq \min\{2, |y|/a\}$. Because $\mu(\mathbb{R}) = 1$, $[\mu * \rho_a](x) - \rho_a(x) = \int [\delta_y * \rho_a](x) - \rho_a(x)\, d\mu(y)$, and so

$$
\|\mu * \rho_a - \rho_a\| \leq \int_{\mathbb{R}} \|\delta_y * \rho_a - \rho_a\|\, d|\mu|(y) \leq \int_{\mathbb{R}} \min\{2, |y|/a\}\, d|\mu|(y),
$$

which tends to 0 as a tends to infinity.

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

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

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 $\nu = \rho_a^{(2)} * \tau \in \mathcal{P}_n$. $\qquad\square$

It is important that the proof of this lemma is constructive; indeed, we get that

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

for c that can be arbitrarily small, and a that may need to be correspondingly large. Since ρ_a is the distribution of a Gaussian with zero mean and standard deviation a, ν is the distribution of

$$Z = W + (X_0 + X_1 + \cdots + X_N),$$

where W is a Gaussian with mean zero and standard deviation $\sqrt{2}a$, $X_0 = 0$, $X_1, X_2, \ldots$ are i.i.d. random variables with distribution π (defined in the proof), and N is geometric with parameter c. In particular, when c is small, Z is close in total variation to W, as $\|\nu - \rho_a^{(2)}\| \leq c$. This shows Remark 1, taking $c = \varepsilon$.

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

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

for every $A \subseteq \mathbb{R}$. By Tonelli's Theorem

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

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

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

17

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),

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

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 $\frac{1}{k}(F_Y - F_X)$. Therefore, $\sigma * \eta$ is absolutely continuous as well, and its density s satisfies, almost everywhere,[15]

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

Because $\sigma * \eta$ is a probability measure and $d > 0$, then $F_{Y+Z}(x) \ge F_{X+Z}(x)$ for almost every x. Since the cdfs are right-continuous, this implies $F_{Y+Z} \ge F_{X+Z}$. Furthermore, this inequality is strict somewhere, since the integral of $k(F_{Y+Z} - F_{X+Z})$ 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 $\mathbb{E}[X] \le \mathbb{E}[Y]$ then there does not exist a random variable Z such that $X + Z >_1 Y + Z$. This is immediate if Z has finite expectation, since, as is well known, $X + Z >_1 Y + Z$ would imply that $\mathbb{E}[X + Z] > \mathbb{E}[Y + Z]$. For general Z, note that the measure

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

satisfies $\sigma'(\mathbb{R}) \le 0$, and so for any probability measure η it holds that $[\sigma' * \eta](\mathbb{R}) \le 0$. 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 $k = \frac{1}{2}(\mathrm{Var}[Y] - \mathrm{Var}[X]) > 0$. Let F_X and F_Y be the cumulative distribution functions of X and Y. Define the signed measure σ as

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

---

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

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 $\int_{\mathbb{R}} |\int_{-\infty}^{t} F_X(u) - F_Y(u) \, \mathrm{d}u| \, \mathrm{d}t < \infty$. See, for instance, (Rachev et al., 2013, Lemma 15.2.1) or (Rachev and Stoyanov, 2008, p.160). Hence σ is bounded. We show that a calculation similar to the one used in Theorem 1 shows that $\sigma(\mathbb{R}) = 1$ and that $\sigma \in \mathcal{M}_{n-2}$. More generally, we claim that for every k,

$$\int_{-\infty}^{\infty} t^k \mathrm{d}\sigma(t) = \frac{1}{(k+1)(k+2)} \int_{-\infty}^{\infty} u^{k+2} \, \mathrm{d}(F_Y - F_X)(u).$$

As is well known, integration by parts implies

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

Hence, for every n and k,

$$
\begin{aligned}
\int_{-n}^{n} t^k \int_{-\infty}^{t} F_Y(u) - F_X(u) \, \mathrm{d}u \, \mathrm{d}t &= \int_{-n}^{n} t^k (\mathbb{E}[(t - Y)^+] - \mathbb{E}[(t - X)^+]) \, \mathrm{d}t. \\
&= \int_{-n}^{n} t^k \int_{-\infty}^{n} (t - u)^+ \, \mathrm{d}(F_Y - F_X)(u) \, \mathrm{d}t. \\
&= \int_{-\infty}^{n} \int_{-n}^{n} t^k (t - u)^+ \, \mathrm{d}t \, \mathrm{d}(F_Y - F_X)(u) \\
&= \int_{-\infty}^{n} \int_{u}^{n} t^k (t - u) \, \mathrm{d}t \, \mathrm{d}(F_Y - F_X)(u) \\
&= \int_{-\infty}^{n} \left(\frac{t^{k+2}}{k+2} - u \frac{t^{k+1}}{k+1} \right) \Big|_{u}^{n} \, \mathrm{d}(F_Y - F_X)(u) \\
&= \int_{-\infty}^{n} \frac{n^{k+2}}{k+2} - u \frac{n^{k+1}}{k+1} + \frac{u^{k+2}}{(k+1)(k+2)} \, \mathrm{d}(F_Y - F_X)(u)
\end{aligned}
$$

For every n, the last integral is well defined and finite provided $X, Y \in \mathcal{P}_{k+2}$. It is a standard result that every $W \in \mathcal{P}_{k+2}$ satisfies $n^{k+2}[1 - F_W(n) + F_W(-n)] \to 0$ as $n \to \infty$.[16] Hence $n^{k+2}(F_Y - F_X)(n) \to 0$ as $n \to \infty$. To see that $n^{k+1} \int_{-\infty}^{n} u \, \mathrm{d}(F_Y - F_X)(u)$ converges to 0, notice that

$$\int_{-\infty}^{n} n^{k+1} u \, \mathrm{d}(F_Y - F_X)(u) \leq \int_{-n}^{n} n^{k+2} \, \mathrm{d}(F_Y + F_X)(u) + \int_{-\infty}^{-n} |u|^{k+2} \, \mathrm{d}(F_Y + F_X)(u).$$

The first term on the right hand side converges to 0 by the same argument as above, and the second term converges to 0 from the assumption that $X, Y \in \mathcal{P}_{n+2}$. This concludes the proof of the claim.

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

---

[16]This follows from

$$n^{k+2}(\mathbb{P}[|W|^{k+2} \geq n^{k+2}]) \leq \mathbb{E}[\mathbb{1}_{\{|W|^{k+2} \geq n^{k+2}\}} |W|^{k+2}]$$

and the observation that since W^{k+2} is integrable, the right hand side must converge to 0 as $n \to \infty$.

independent from X and Y with distribution η. Then $s(x)$, the probability density function of $\sigma * \eta$, is

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

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 $\int_{-\infty}^{x} F_{Y+Z}(u) - F_{X+Z}(u) \, du \geq 0$ for all x is a well known condition for second-order stochastic dominance (see Theorem 4.A.2 in Shaked and Shanthikumar, 2007).

C 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 closed-form 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

$$
Z = R_1 + R_2 + (U_1 + U_2 + \cdots U_N). \tag{8}
$$

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 (8), 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.

C.1 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

$$m_a(x) = \frac{1}{2a}\mu([x - a, x + a]) - r_a(x) \tag{9}$$

and which satisfies the following properties:

Lemma 2. *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 $m_a(x) = 0$.

$\square$

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 2 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

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

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

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 (8) in the main text. Finally, the same argument used in the proof of Theorem 1 shows that $X + Z >_1 Y + Z$.

D Calculations For Section 2.2

Let X be a gamble paying $g > 0$ and $-l < 0$ with equal probability. Let $Y = 0$. Following the proof Theorem 1, we first define μ to be the measure with support $[-l, g]$ and density

$$\frac{1}{\mathbb{E}[X]}(F_Y - F_X)(s) = \begin{cases} -\frac{1}{g-l} & s \in [-l, 0] \\ \frac{1}{g-l} & s \in [0, g] \end{cases}$$

and then seek to find a probability measure π, and parameters a and c such that

$$(\rho_a * \mu - \rho_a) + \frac{c}{\sqrt{2}}\pi \geq 0. \tag{10}$$

Let ϕ_a and Φ_a be, respectively, the pdf and CDF of a Gaussian random variable with mean zero and standard deviation a. Then $\rho_a * \mu - \rho_a$ has density

$$t(s) = \frac{1}{g - l}(2\Phi_a(s) - \Phi_a(s - g) - \Phi_a(s + l)) - \phi_a(s).$$

We claim that $t(s) \geq 0$ for $s > g$ and $t(s) \leq 0$ for $s < -l$. By the mean-value theorem, for each s there exist $s_1 \in [s - g, s]$ and $s_2 \in [s, s + l]$ such that $\Phi_a(s) - \Phi_a(s - g) = \phi_a(s_1)g$ and $\Phi_a(s) - \Phi_a(s + l) = \phi_a(s_2)(-l)$. Hence

$$t(s) = \frac{1}{g - l}(\phi_a(s_1)g - \phi_a(s_2)l) - \phi_a(s). \tag{11}$$

Suppose $s - g > 0$. Then $\phi_a(s_1)g > \phi_a(s)g$. Hence (11) is greater than

$$\frac{1}{g - l}(\phi_a(s)g - \phi_a(s_2)l) - \phi_a(s) = \frac{l}{g - l}(\phi_a(s) - \phi_a(s_2)).$$

g	l	σ_W	c	σ_Z
12	10	3,500	0.014	3,525
12	10	4,000	0.011	4,025
100	50	4,500	0.022	4,551
100	50	10,000	0.010	10,050
100	70	11,000	0.018	11,102

Table 1: Properties of the noise $Z = W + U_1 + \ldots + U_N$, where W is Gaussian with mean zero and standard deviation σ_W, (U_i) are i.i.d. random variables, N is geometrically distributed with parameter c, and σ_Z is the standard deviation of Z.

Since $0 < s < s_2$ then $\phi_a(s) - \phi_a(s_2) > 0$. This shows that $t(s) \geq 0$. Now suppose $s < -l$. Then $\phi_a(s_1)g < \phi_a(s)g$. Hence (11) is lower than

$$\frac{1}{g-l}(\phi_a(s)g - \phi_a(s_2)l) - \phi_a(s) = \frac{l}{g-l}(\phi_a(s) - \phi_a(s_2)),$$

and because $s < s_2 < 0$ then $\phi_a(s) - \phi_a(s_2) < 0$. Therefore $t(s) \leq 0$. This concludes the proof of the claim. We take π to be absolutely continuous with density f. Then (10) is satisfied if

$$f(s) \geq -\frac{\sqrt{2}}{c}t(s).$$

We take $f(s) = 0$ if $s > g$ (since then $t(s) \leq 0$) and

$$f(s) = \frac{\sqrt{2}}{c}\,|t(s)| \quad \text{if} \quad s \leq g$$

and then set

$$c = \sqrt{2}\int_{-\infty}^{g}|t(s)|\,\mathrm{d}s$$

so that f integrates to 1.

Given the parameters g, l and a, the coefficient c and the pdf f can be computed numerically. This allows us to describe some quantitative features of the background noise Z, as discussed in section 2.2.

Table 1 provides some examples. For instance, when the binary gamble X pays \$12 and -\$10 dollars with probability $1/2$, there exists a background risk $Z = W + U_1 + \ldots + U_N$ such that $X + Z >_1 Z$ and: the Gaussian W has standard deviation 3,500, $\mathbb{P}[N = n] = (0.014)^n$ for $n \in \{0, 1, 2, \ldots\}$, and Z has standard deviation 3,525.