How risky is a random process?

Sudhir A. Shah

Email: sudhir@econdse.org
Department of Economics
Delhi School of Economics

Working Paper No. 252

CENTRE FOR DEVELOPMENT ECONOMICS
DELHI SCHOOL OF ECONOMICS
DELHI 110007

How risky is a random process?

Sudhir A. Shah[*]

January 13, 2016

Abstract

The riskiness of random processes is compared by (a) employing a decision-theoretic equivalence between processes and lotteries on path-spaces to identify the riskiness of the former with that of the latter, and (b) using the theory of comparative riskiness of lotteries over vector spaces to compare the riskiness of lotteries on a given path-space. We derive the equivalence used in step (a) and contribute a new criterion to the theory applied in step (b). The new criterion, involving a generalized form of second order stochastic dominance, is shown to be valid by establishing its equivalence to the standard decision-theoretic criterion. We demonstrate its tractability *via* diverse economic applications featuring risk embodied in random processes.

JEL classification: C02, D01, D63, D81

Key words: random processes, vector outcomes, comparative riskiness, generalized second order stochastic dominance

1 Introduction

The answer to the above question relies on the theory of comparative riskiness of lotteries. The cornerstone of this theory is the decision-theoretic criterion (henceforth, DTC): one lottery over a set of outcomes is said to be riskier than another one if every decision-maker with an admissible concave (Bernoulli-von Neumann-Morgenstern) utility defined on the outcomes derives at least as much expected utility from the latter as from the former.[1]

The comparative riskiness of random variables can be assessed by identifying it with the comparative riskiness of their distributions on the real line and assessing the latter using DTC or its characterizations in Rothschild and Stiglitz [22]. This is a legitimate strategy because a decision-maker's expected utility is invariant with respect to the implied change of integrating variable (cf. equation (1)).

---

[*]Department of Economics, Delhi School of Economics, University of Delhi, Delhi 110007, India. Email: sudhir@econdse.org
[1]A concave utility is an abstraction of a risk averse utility (Arrow [1], Pratt [21]); see Section 5.1 for the relationship between these notions in the context of this paper.

The proliferation of economic models with risk embodied in random processes (henceforth, processes) motivates the development of a theory for comparing the riskiness of processes, which can be used to study the effects of such risks on the endogenous variables of these models.

While the outcomes of a random variable are real numbers, the outcomes of a process, namely its sample paths, are real-valued (possibly, even vector-valued) functions over a time domain. Since these functions belong to a path-space, which is a subset of a vector space, the strategy for comparing the riskiness of processes must suitably generalize the strategy for comparing the riskiness of random variables.

The first step of the strategy is to establish a duality, i.e., a decision-theoretic equivalence, between processes and lotteries on path-spaces, so that the riskiness of a process can be legitimately identified with that of the lottery dual to it. This is done in Section 3. This duality allows seamless translation between processes and lotteries on path-spaces.

Since we regard DTC as the touchstone for a valid comparison of riskiness, the second step is to compare the riskiness of the dual lotteries on a path-space by using DTC or a criterion equivalent to it. Characterizations of DTC are important, if only because the tractability of different, albeit equivalent, criteria can vary with the applications. As most commonly used processes have infinite time domains, their path-spaces are embedded in various infinite-dimensional vector spaces. Therefore, a useful characterization of DTC must hold for as general a class of infinite-dimensional vector outcome spaces as possible. What is known about such characterizations?

The economics literature reveals no progress towards an answer since the Euclidean space results of Russell and Seo [23]. The other source of answers is the statistical literature on convex orders (surveyed in Mosler and Scarsini [17], and more recently in Shaked and Shanthikumar [25], Sections 3.A, 3.E, 7.A and 7.E) and comparison of experiments (surveyed in Le Cam [16] and Torgersen [29]). This literature yields the two classical characterizations that are cited in Section 5.2 and stated in Appendix A. Neither DTC nor its classical characterizations have found traction in economic applications featuring random infinite-dimensional outcomes.

So, it is necessary to find a criterion that is equivalent to DTC in infinite-dimensional settings and tractable in economic applications. Our **main result**, Theorem 6.3, provides a criterion that meets both requirements. This result characterizes DTC in terms of second order stochastic dominance applied to the utility distributions generated by the given lotteries over vector outcomes and the admissible utilities defined over these outcomes. Given *any real vector space*, the characterization holds for every measurable convex outcome space embedded in that vector space and every family of bounded, concave and measurable utilities on the outcomes that satisfies an easily verifiable closure property.

Special versions of this result require topological or ordering structure

on the outcome space if the admissible utilities are to have continuity or monotonicity properties (Corollary 6.4). Since all the admissible families of utilities appearing in this literature satisfy the specified closure property, the result characterizes many familiar versions of DTC in a unified manner.

In addition to our main result, Theorems 6.5-6.7 show that the vector outcomes theory generalizes various key features of the Rothschild-Stiglitz theory; for instance, comparisons of vector riskiness generate a partial ordering of the risks, imply mean-preservation and yield a comparison of the risk premia. While the scalar versions of these results are familiar, the vector versions are new, not merely because they hold for substantially more general settings, but also because these settings force conceptual innovations; see a synopsis of these results in Section 5.3 and details in Sections 6.2-6.4.

When faced with a problem involving processes, it is tempting to try to finesse it by pathwise aggregating the sample paths, thereby deriving an aggregate random variable from each process. Thereafter, one might identify the comparative riskiness of processes with that of the corresponding aggregate random variables, where the latter comparison uses the Rothschild-Stiglitz theory. Theorem 6.8 shows that this ploy is illegitimate: for, arguably, the largest class of vector outcome spaces that appear in economic models, the ordering of the riskiness of vector outcome lotteries using DTC is *necessarily incongruent* with the ordering of the riskiness of aggregate outcome distributions using the Rothschild-Stiglitz theory.

Since Theorems 6.3-6.8 hold for lotteries over appropriate vector outcome spaces, they necessarily hold for lotteries over appropriate path-spaces. Using the duality result, the latter versions of these theorems are readily translated into propositions regarding processes.

Finally, since our aim was to characterize DTC so as to facilitate applications, we demonstrate the tractability of the new criterion by using it to compare riskiness in economic applications where risk is embodied in processes. A typical economic application of the theory of comparative riskiness examines how variations in exogenously specified riskiness affect the choices made by a decision-maker seeking to maximize expected utility. Such decision-makers care about the riskiness of the primitive outcomes, namely the sample paths, only insofar as it affects the relationship between their decision and the distribution of the derived outcomes, namely the utilities. Using the classical criteria in these applications requires the translation of criteria based purely on the primitive outcomes into criteria based on the derived outcomes. Instead, our criterion is easier to use in these problems as it is based directly on the distributions of the derived outcomes. We substantiate this claim in Section 7, where we use our criterion to identify conditions under which the following results hold:

1. A bidder in a sealed-bid (first-price or second-price) auction will bid more conservatively with greater riskiness of the random process de-

termining his valuation of the prize.

2. A regulated utility, which is penalized if it fails to meet a random demand flow over a period, will choose a higher capacity in response to greater riskiness of the demand process.

3. A firm's manager, who is incentivized to minimize stock-outs of its product, will choose a larger production capacity to deal with greater riskiness of the random orders process.

4. A risk averse investor will increase the risk-free component of his portfolio in response to increased riskiness of the stream of returns generated by the risky asset.

5. A planner will provide more of a public good in response to greater riskiness of the consumption allocations of the citizens.

6. The principal's contribution in a team moral hazard situation increases with greater riskiness of the effort levels chosen by the agents.

Such problems have hitherto been modeled with real-outcome risks, but this surely was a concession to the available tool-kit, rather than an attempt at verisimilitude. At one level, our results simply confirm intuitions derived from the real outcomes theory. At another level, they use appropriate models and tools to convert conjectures into theorems.

All the applications are modeled with risk embodied in a process as the sole complication and other features pared down to be as elementary as possible without sacrificing economic substance. For instance, all the applications feature a real-valued decision variable. Applications featuring multi-dimensional controls or game-theoretic considerations are beyond the scope of this paper as their complexity demands longer dedicated treatments.

The contents of the sections not discussed above are as follows. Section 2 states our technical conventions. Section 4 recalls the Rothschild-Stiglitz theory. Section 5 states the formalism underlying the vector outcomes theory and uses it to catalog the relevant literature and our contributions to it. We conclude in Section 8. Appendix A states the classical equivalence results. Proofs of all results other than the main result are collected in Appendix B.

2 Notation and conventions

The following conventions shall apply throughout this paper.

$\neg P$ denotes the negation of proposition P. The subset relation $\subset$ is used in the weak sense. $\mathcal{N}$ is the set of natural numbers. The real line $\Re$ is given the Euclidean metric topology. A product of topological spaces is given the product topology and a subset of a topological space is given the subspace

topology. A topological space A is given the Borel σ-algebra $\mathcal{B}(A)$. For a topological space A, the set of continuous functions $f : A \to \Re$ is denoted by $\mathcal{C}(A)$. $\Delta(A)$ denotes the set of lotteries, i.e., countably additive probability measures, on a measurable space $(A, \mathcal{A})$. The distribution function corresponding to a lottery $\nu \in \Delta(\Re)$ is given by $F(., \nu) = \nu(-\infty, .]$.

Every vector space is over the field $\Re$. Let $(X, \geq)$ be an ordered vector space. For $x, y \in X$, we say $x > y$ if $x \geq y$ and $\neg y \geq x$. For $O \subset X$, we say $f : O \to \Re$ is increasing (resp. strictly increasing, decreasing) if $x, y \in O$ and $x > y$ implies $f(x) \geq f(y)$ (resp. $f(x) > f(y)$, $f(x) \leq f(y)$).

Consider $f : A \to \Re$ with $A \subset \Re$. f is said to be $\mathcal{C}^k$ if it is k-times continuously differentiable for $k \in \mathcal{N} \cup \{\infty\}$. For $k \in \mathcal{N}$, $D^k f(x)$ is the k-th derivative of f at $x \in A$. $f(x+)$ (resp. $f(x-)$) is the right-hand (resp. left-hand) limit of f at $x \in A$.

3 Duality of processes and lotteries

In this section, we show that a process is dual to a lottery on its path-space and the expected utility derived from a process is identical to the expected utility derived from its dual lottery. So, the problem of comparing the riskiness of processes is decision-theoretically equivalent to the problem of comparing the riskiness of their dual lotteries using DTC.

Given $T \subset \Re$, let $\Re^T$ be the set of functions $x : T \to \Re$. $\Re^T$ is a vector space given $(x + y)(.) := x(.) + y(.)$ and $(rx)(.) := rx(.)$ for $x, y \in \Re^T$ and $r \in \Re$. A measurable space $(O, \mathcal{O})$ is called a *path-space* if $O \subset \Re^T$.[2]

Definition 3.1 *A random process is a collection $\mathcal{Y} = (\Omega, \mathcal{F}, P; T, y)$, where $(\Omega, \mathcal{F}, P)$ is a probability space, $T \subset \Re$ is the time domain and the function $y : T \times \Omega \to \Re$ is such that $y(t, .)$ is measurable for every $t \in T$. Each $\omega \in \Omega$ generates the sample path $\hat{y}(\omega) \in \Re^T$, given by $\hat{y}(\omega)(.) := y(., \omega)$.*

The next definition links processes to lotteries on path-spaces by requiring $\hat{y}$ to be measurable, i.e., by requiring it to be a random element, which is the natural generalization of the notion of a random variable.

Definition 3.2 *A process $\mathcal{Y} = (\Omega, \mathcal{F}, P; T, y)$ is said to be admissible with respect to a path-space $(O, \mathcal{O})$ if $\hat{y}(\Omega) \subset O$ and $\hat{y} : \Omega \to O$ is measurable.[3] The image measure $P \circ \hat{y}^{-1} \in \Delta(O)$ is called the distribution generated by an admissible process $\mathcal{Y}$ on $(O, \mathcal{O})$.*

---

[2]Consider a more general path-space $O \subset Y^T$ where Y is a vector space. Since Y^T is a vector space, our main result will apply to O. However, we do not pursue this generality and restrict attention to $Y = \Re$ for simplicity and concreteness.

[3]For instance, it is shown in Shah [24] that various Wiener-related processes and a large class of second order processes are admissible with respect to appropriate path-spaces.

Given a path-space, we declare processes to be equivalent if they are admissible with respect to it and generate the same distribution on it. Then, the mapping of the resulting equivalence classes of processes to their respective distributions on the path-space is automatically injective. We now show that, for an appropriate path-space, this mapping is also surjective.

Definition 3.3 *Consider a path-space $(O, \mathcal{O})$. Let $e : T \times O \to \Re$ be the evaluation function defined by $e(t, x) = x(t)$. We say that $(O, \mathcal{O})$ is admissible if $e(t, .)$ is measurable for every $t \in T$.*

Theorem 3.4 *If a path-space $(O, \mathcal{O})$ is admissible and $P \in \Delta(O)$, then P is the distribution of a process that is admissible with respect to $(O, \mathcal{O})$.*

So, if $(O, \mathcal{O})$ is an admissible path-space, then the mapping of equivalence classes of admissible processes to their respective distributions in $\Delta(O)$ is bijective. Moreover, an admissible process is decision-theoretically equivalent to its distribution on the path-space.

For the last claim, consider a process $\mathcal{Y}$ that is admissible with respect to a path-space $(O, \mathcal{O})$. If $u : O \to \Re$ is an integrable utility, then the change of variable formula (Billingsley [4], Theorem 16.13) implies

$$\int_{\Omega} P(d\omega) \, u \circ \hat{y}(\omega) = \int_{O} P \circ \hat{y}^{-1}(dx) \, u(x) \tag{1}$$

i.e., expected utility is invariant across the dual representations of risk, where the left-hand side and the right-hand side of (1) are the expected utilities derived from the process and its distribution respectively.

The above-described decision-theoretic duality justifies the identification of the riskiness of a process with the riskiness of its distribution.

Definition 3.5 *Suppose $(O, \mathcal{O})$ is a path-space and $\succeq$ is a relation that compares the riskiness of lotteries in $\Delta(O)$. If processes $\mathcal{Y}_1 = (\Omega_1, \mathcal{F}_1, P_1; T, y_1)$ and $\mathcal{Y}_2 = (\Omega_2, \mathcal{F}_2, P_2; T, y_2)$ are admissible with respect to $(O, \mathcal{O})$, then we say that $\mathcal{Y}_1$ is riskier than $\mathcal{Y}_2$ if $P_1 \circ \hat{y}_1^{-1} \succeq P_2 \circ \hat{y}_2^{-1}$.*

4 Real outcomes theory

As the real outcomes theory is the model for the vector outcomes theory of comparative riskiness, the latter seeks to emulate the key features of the former. In this section, we note these congruences and use them to interpret some modeling choices made in the more general theory.

Consider a convex set $O \subset \Re$ and lotteries $\mu, \lambda \in \Delta(O)$. We say $\mu R_1 \lambda$ if $\int_O \mu(dx) \, v(x) \leq \int_O \lambda(dx) \, v(x)$ for every function $v : O \to \Re$ that is bounded, measurable and concave. R_1 is a special version of DTC (cf. Definition 5.3).

We say $\mu R_2 \lambda$ if μ and λ have the same mean and λ second order stochastically dominates μ, i.e., $\int_{(-\infty,y]} dx\,[F(x,\mu) - F(x,\lambda)] \geq 0$ for every $y \in \Re$.

We say $\mu R_3 \lambda$ if λ and μ are the distributions of random variables f and g respectively and the expectation of g conditional on f is equal to f almost surely (cf. Definitions A.3 and A.4).

The main result of the Rothschild-Stiglitz theory is: if O is a convex and compact subset of $\Re$, then R_1, R_2 and R_3 are identical partial orders (Rothschild and Stiglitz [22], Theorem 2). Reflexivity implies that "riskier than" is meant in the weak sense. Incompleteness means that this theory does not rationalize any cardinal measure of riskiness. The partial order property is generalized to the vector outcomes setting in Theorem 6.5 and the equivalence of R_1 and R_3 is generalized in Theorem A.5.

The definitions of R_1 and R_3 do not call for the existence or knowledge of the means of μ and λ (cf. Definitions 5.3, A.3 and A.4), much less their equality. Nevertheless, as is well-known, $\mu R_1 \lambda$ (resp. $\mu R_3 \lambda$) implies that μ and λ have the same mean (cf. Theorem 6.6).[4]

It is also known that the risk premium attached to a lottery by a strictly increasing, concave and continuous utility increases with the riskiness of the lottery (cf. Theorem 6.7).

There is no presumption in R_1 that an admissible v must be monotonic; Equation (2) and Footnote 3 in Rothschild and Stiglitz [22] affirm this explicitly (cf. Definition 5.3). If the outcomes are known to be agreeable (resp. disagreeable) in a particular context, say monetary gains (resp. pain, prison terms, pollution), then it is sensible for v to be increasing (resp. decreasing) in that context. Or, a risk averse Goldilocks may have a concave utility that increases to a peak and then decreases. The absence of a monotonicity postulate admits many interpretations of the outcomes and the preferences. This agnosticism makes for a pure theory of comparative riskiness as its interpretation is independent of any particular context.

Insisting that the utilities underlying R_1 must be increasing restricts the theory's domain to risks that are sensibly reducible to "monetary" lotteries. The monotonicity requirement also confounds a decision-maker's attitude to risk with a preference for "more". To see this, consider lottery δ_0 (resp. δ_1) that gives prize 0 (resp. 1) almost surely. Every increasing utility, whether concave or otherwise, prefers δ_1 to δ_0. So, if the definition of R_1 is amended to require that an admissible v must be increasing, in addition to the other requirements, then the amended criterion will declare δ_0 to be riskier than δ_1. This unconvincing conclusion is driven purely by monotonicity and has nothing to do with risk aversion or risk.

---

[4]The means exist as O is compact. To check their equality given $\mu R_1 \lambda$, set $v = I$ and $v = -I$, where I is the identity mapping on O. Given $\mu R_3 \lambda$ and the underlying probability space $(\Omega, \mathcal{F}, P)$, let E be the expectation operator and let $E_f g$ be the conditional expectation of g given f. Then, $\int_O \mu(dx)\,x = \int_O P \circ g^{-1}(dx)\,x = \int_\Omega P(d\omega)\,g(\omega) = Eg$. Similarly, $\int_O \lambda(dx)\,x = Ef$. Therefore, $\int_O \lambda(dx)\,x = Ef = EE_f g = Eg = \int_O \mu(dx)\,x$.

In order to eliminate such anomalies, every version of R_1 that insists on increasing utilities also *imposes* mean-preservation as a prerequisite for comparable riskiness. This supplementary condition renders non-comparable the riskiness of δ_0 and δ_1. R_1 reaches the same conclusion by admitting *all* concave utilities. Since admissible increasing and decreasing utilities imply contradictory assessments of δ_0 and δ_1, their riskiness is non-comparable.

5 Vector outcomes theory

This section provides a formal synopsis of the theory of comparative riskiness of lotteries with vector outcomes.

5.1 Definitions

Definition 5.1 *An outcome space is a pair (X, O), where O is a set of outcomes in a real vector space X. The theory uses the following specifications.*

(1) O is a measurable convex subset of X.

(2) O is a convex subset of a topological vector space X.

(3) O is a convex, compact and metrizable subset of a locally convex topological vector space X.

(4) O is a convex and compact subset of a separable Banach space X.

(5) O is a convex and compact subset of a Euclidean space X.

If an outcome space (X, O) satisfies specification (n) for $n \in \{1, \dots, 5\}$, then we say (X, O) belongs to Setting n and is admissible.

Given the conventions of Section 2, the settings are nested as follows:

$$\text{Setting } n \supset \text{Setting } n+1, \quad n \in \{1, \dots, 4\} \tag{2}$$

So, if a definition or a proposition is valid for outcome spaces in Setting $n \in \{1, \dots, 4\}$, then it is also valid for outcome spaces in Setting $n + 1$.

A subset of Setting n is specified by requiring X in that setting to be an ordered vector space. The resulting ordered settings satisfy a nesting scheme analogous to (2).

Consider an admissible outcome space (X, O). Let V be the set of bounded, concave and measurable functions $v : O \to \Re$. As $v \in V$ is bounded, $v_0 := \inf v(O) \in \Re$ and $v_1 := \sup v(O) \in \Re$. Moreover, every $v \in V$ is integrable with respect to every $\mu \in \Delta(O)$. If X is topological, then let V_c (resp. V_{sc}) be the set of continuous (resp. upper semicontinuous) functions in V. If X is ordered, then let V^m (resp. V_c^m, V_{sc}^m) be the set of increasing functions in V (resp. V_c, V_{sc}).

Definition 5.2 *Consider an admissible outcome space (X, O). $V' \subset V$ is said to be an admissible set of utilities with respect to (X, O) if it is well-defined, given the specified properties of (X, O), i.e., it is possible to determine for every $v \in V$, whether $v \in V'$.*

8

For instance, if (X, O) is an admissible outcome space without topological structure, then V is admissible with respect to (X, O), but V_c and V_{sc} are not admissible.

Definition 5.3 *Consider an admissible outcome space (X, O). Let $V' \subset V$ be a set of utilities that is admissible with respect to (X, O). Given $\mu, \lambda \in \Delta(O)$, we say*

(1) $\mu \succeq_1^{V'} \lambda$ if $\int_O \mu(dx) v(x) \leq \int_O \lambda(dx) v(x)$ for every $v \in V'$, and

(2) $\mu \succeq_2^{V'} \lambda$ if $\int_{(-\infty, y]} dx \left[F(x, \mu \circ v^{-1}) - F(x, \lambda \circ v^{-1}) \right] \geq 0$ for all $v \in V'$ and $y \in \Re$.

Relation $\succeq_1^{V'}$ is a version of DTC. The statement $\mu \succeq_2^{V'} \lambda$ means that the utility distribution $\lambda \circ v^{-1}$ second order stochastically dominates the utility distribution $\mu \circ v^{-1}$ for every admissible utility $v \in V'$.

We end this section by relating concavity and risk aversion in the context of the above formalism. Given an admissible outcome space (X, O), consider a bounded and measurable utility $v : O \to \Re$. Utility v is said to be *risk averse* if $v(m_\mu) \geq \int_O \mu(dx) v(x)$ for every $\mu \in \Delta(O)$, where $m_\mu \in O$ denotes "the mean of μ". While $\int_O \mu(dx) v(x)$ is well-defined and finite even in Setting 1, verification of risk aversion requires the definition and existence of a unique $m_\mu \in O$.

If (X, O) belongs to Setting 5 with $X = \Re^n$ and $m_\mu := \left(\int_O \mu(dx) \pi_i(x) \right)_{i=1}^n$ where $\pi_i : \Re^n \to \Re$ is the i-th projection mapping, then m_μ has the requisite properties and v is risk averse if and only if $v \in V$ (Perlman [19], Proposition 1.1). More generally, our results in Section 6.3 point to the Pettis (or weak) mean as the appropriate definition of m_μ in Setting 3. In this setting, if m_μ is the Pettis mean and v is upper semicontinuous, then m_μ has the requisite properties (Theorem 6.6) and v is risk averse if and only if $v \in V_{sc}$ (Perlman [19], Theorem 3.10). Settings 1 and 2 lack the structure to ensure that m_μ has the requisite properties. Therefore, in these general settings, concavity serves as an abstract representation of risk aversion.

5.2 The known results

Theorem A.2 (Blackwell [5], [6], Cartier, Fell and Meyer [7], Hardy, Littlewood and Pólya [12], Sherman [26], Stein [27], Strassen [28]) in Appendix A shows that $\succeq_1^{V_c} = \succeq_3$ for outcome spaces in Setting 3, where $\succeq_3$ refers to the existence of an appropriate dilatation on O.

Theorem A.5 (Blackwell [5], [6], Sherman [26], Stein [27], Strassen [28]) in Appendix A shows that $\succeq_1^{V_c} = \succeq_4$ for outcome spaces in Setting 4, where $\succeq_4$ is the vector outcome generalization of R_3.

It is also known that $\succeq_1^{V_{sc}^m} = \succeq_2^{V_{sc}^m}$ for outcome spaces in Setting 5 (Russell and Seo [23], Remark 1). However, this result has two limitations.

First, as noted in Section 4, the monotonicity condition on the utilities in V_{sc}^m means that $\succeq_1^{V_{sc}^m}$ and $\succeq_2^{V_{sc}^m}$ cannot be interpreted as comparing riskiness

unless their definitions are strengthened with supplementary conditions, say "mean-preservation". But, this raises the question: what is an appropriate definition of the "mean" of a vector outcome lottery?

Second, even if supplementary restrictions modify $\succeq_1^{V_{sc}^m}$ and $\succeq_2^{V_{sc}^m}$ into valid measures of riskiness, the restriction of the result to Setting 5 means that it cannot be used to compare the riskiness of processes with infinite time domains, which is our primary motivation.

Our results will resolve both these issues.

5.3 Summary of our theoretical results

Our main result is the omnibus Theorem 6.3: given an admissible setting (X, O), if a set of functions $V' \subset V$ is admissible with respect to (X, O) and satisfies the "closure property", then $\succeq_1^{V'} = \succeq_2^{V'}$. Corollary 6.4 interprets V' in this equality as various familiar admissible sets of utilities.

For outcome spaces in Setting 1, the equality holds for $V' = V$. While relation $\succeq_1^V$ is the broadest version of DTC and the *verbatim* extension of R_1, relation $\succeq_2^V$ generalizes the Russell-Seo relation $\succeq_2^{V_{sc}^m}$. This result characterizes DTC in the most general possible setting.

For outcome spaces in Setting 2, the equality holds for $V' \in \{V_c, V_{sc}\}$.

Using (2) and setting $V' = V_c$, $\succeq_1^{V_c}$ and $\succeq_2^{V_c}$ are equal to $\succeq_3$ (resp. $\succeq_4$) for outcome spaces in Setting 3 (resp. Setting 4). This extends the classical characterizations in Theorems A.2 and A.5.

For all the above-mentioned settings and specifications of V', the relations $\succeq_1^{V'}$ and $\succeq_2^{V'}$ are equivalent ways to compare the riskiness of lotteries.

Given settings with ordering structure, the equality also holds for $V' \in \{V^m, V_c^m, V_{sc}^m\}$. For instance, if (X, O) belongs to Setting 1 and X is ordered, then $\succeq_1^{V^m} = \succeq_2^{V^m}$. Moreover, if (X, O) belongs to Setting 2 and X is ordered, then $\succeq_1^{V_c^m} = \succeq_2^{V_c^m}$ and $\succeq_1^{V_{sc}^m} = \succeq_2^{V_{sc}^m}$, which substantially generalizes the Russell-Seo result. *For ordered settings, the relations based on monotonic preferences cannot be interpreted as comparisons of riskiness, unless their definitions are supplemented by Pettis mean-preservation.*

We also derive the following three results that generalize the corresponding results flowing from the Rothschild-Stiglitz theory.

For an outcome space (X, O) in Setting 1, Theorem 6.5 shows that $\succeq_1^V$ imposes a partial order on the quotient set of equivalence classes generated by a natural notion of identically risky lotteries in $\Delta(O)$. This generalizes the Rothschild-Stiglitz result that R_1 imposes a partial order on $\Delta(O)$.

For an outcome space in Setting 3, Theorem 6.6 shows that each lottery must have a unique Pettis mean, and if one lottery is riskier than another in terms of $\succeq_1^{V_c}$, then their Pettis means must be the same. So, Pettis mean-preservation is a natural consequence of purely decision-theoretic considerations when monotonicity is not imposed on admissible preferences. This also suggests that the Pettis mean is the appropriate notion of the

10

"mean" if mean-preservation is to be *imposed* as a supplementary restriction in vector outcome settings. This result generalizes the Rothschild-Stiglitz mean-preservation property derived in Footnote 4.

For an outcome space (X, O) in Setting 3, if X is ordered and one lottery is riskier than another, then Theorem 6.7 shows that, for every continuous, concave and strictly increasing utility, the set of risk premia associated with the riskier lottery must be "greater than" the set of risk premia associated with the less risky lottery.

We show in Theorem 6.8 that the need for a vector outcomes theory of comparative riskiness cannot be finessed by using the real outcomes theory to compare random aggregate outcomes. This issue is peculiar to the vector outcomes theory and cannot arise in the real outcomes theory.

6 Our theoretical results

6.1 Equivalence

Let U be the set of functions $u(., y) : \Re \to \Re$, where $y \in \Re$ and $u(., y) = \min\{., y\}$. Each $u \in U$ is increasing, concave, continuous and measurable.

Let (X, O) be an admissible setting and let $V' \subset V$ be a set of functions that is admissible with respect to (X, O). V' is said to be closed under U if $u \circ v \in V'$ for all $u \in U$ and $v \in V'$.[5]

Remark 6.1 *If (X, O) is an outcome space in Setting 1, then V is closed under U; if X is an ordered vector space, then V^m is closed under U. If (X, O) is in Setting 2, then V_c and V_{sc} are closed under U; if X is an ordered vector space, then V_c^m and V_{sc}^m are closed under U.*

Next, we perform an elementary calculation leading to our main result.

Lemma 6.2 *Let $(O, \mathcal{O})$ be a measurable space, $y \in \Re$, $u(., y) \in U$ and $\mu \in \Delta(O)$. If $v : O \to \Re$ is measurable and bounded, then $\int_O \mu(dx) u(v(x), y) = y - \int_{(-\infty, y]} dx F(x, \mu \circ v^{-1})$.*

Proof. By changing variables, $\int_O \mu(dx) u(v(x), y) = \int_\Re \mu \circ v^{-1}(dx) u(x, y) = \int_{(-\infty, v_1]} u(x, y) dF(x, \mu \circ v^{-1})$. As $F(., \mu \circ v^{-1}) = 0$ on $(-\infty, v_0)$ and $F(., \mu \circ v^{-1}) = 1$ on $[v_1, \infty)$, integrating by parts yields $\int_{(-\infty, v_1]} u(x, y) dF(x, \mu \circ v^{-1}) = u(v_1, y) - \int_{(-\infty, v_1]} F(x, \mu \circ v^{-1}) du(x, y)$. Note that

$$u(t, y) - u(s, y) = \begin{cases} t - s, & \text{if } s \le t \le y \\ 0, & \text{if } y \le s \le t \end{cases}$$

---

[5] The following characterization of such a V' might be of interest although we do not use it; see Howie [13] for these concepts. Clearly, $(U, \circ)$ is a semigroup. Each $f \in U$ yields the transformation $T_f : V' \to \Re^O$ by $T_f(v) = f \circ v$. Let $\mathcal{T} = \{T_f \mid f \in U\}$. Given $f, g \in U$, let $T_f * T_g = T_{f \circ g}$. As $(U, \circ)$ is a semigroup, it is easily verified that $(\mathcal{T}, *)$ is a semigroup. Then, $V' \subset V$ is closed under U if and only if $(V', (\mathcal{T}, *))$ is a transformation semigroup.

For $y < v_1$, $u(v_1, y) - \int_{(-\infty, v_1]} F(x, \mu \circ v^{-1}) \, du(x, y) = y - \int_{(-\infty, y]} F(x, \mu \circ v^{-1}) \, du(x, y) - \int_{(y, v_1]} F(x, \mu \circ v^{-1}) \, du(x, y) = y - \int_{(-\infty, y]} dx \, F(x, \mu \circ v^{-1})$.

For $y \geq v_1$, $u(v_1, y) - \int_{(-\infty, v_1]} F(x, \mu \circ v^{-1}) \, du(x, y) = v_1 - \int_{(-\infty, y]} F(x, \mu \circ v^{-1}) \, du(x, y) + \int_{(v_1, y]} F(x, \mu \circ v^{-1}) \, du(x, y) = v_1 - \int_{(-\infty, y]} dx \, F(x, \mu \circ v^{-1}) + y - v_1 = y - \int_{(-\infty, y]} dx \, F(x, \mu \circ v^{-1})$. $\blacksquare$

This yields our main result.

Theorem 6.3 *Consider an admissible setting (X, O) and $V' \subset V$ that is admissible with respect to (X, O). If V' is closed under U, then $\succeq_1^{V'} = \succeq_2^{V'}$.*

Proof. Consider $\mu, \lambda \in \Delta(O)$. Suppose $V' \subset V$ is closed under U.

Let $\mu \succeq_1^{V'} \lambda$, $v \in V'$ and $y \in \Re$. Then, $u(., y) \in U$. As V' is closed under U, $u(v(.), y) \in V'$. Using Lemma 6.2, we have $\int_{(-\infty, y]} dx \, [F(x, \mu \circ v^{-1}) - F(x, \lambda \circ v^{-1})] = \int_O \lambda(dx) \, u(v(x), y) - \int_O \mu(dx) \, u(v(x), y) \geq 0$. So, $\mu \succeq_2^{V'} \lambda$.

Conversely, let $\mu \succeq_2^{V'} \lambda$ and $v \in V'$. Setting $y = v_1$ in Lemma 6.2, $\int_O \lambda(dx) \, v(x) - \int_O \mu(dx) \, v(x) = \int_O \lambda(dx) \, u(v(x), v_1) - \int_O \mu(dx) \, u(v(x), v_1) = \int_{(-\infty, v_1]} dx \, [F(x, \mu \circ v^{-1}) - F(x, \lambda \circ v^{-1})] \geq 0$. Therefore, $\mu \succeq_1^{V'} \lambda$. $\blacksquare$

This result yields a category of concrete equivalence results regarding the various criteria for comparing vector outcome riskiness.

Corollary 6.4 *Consider an outcome space (X, O).*

(A) If it is in Setting 1, then $\succeq_1^V = \succeq_2^V$; if X is an ordered vector space, then we also have $\succeq_1^{V^m} = \succeq_2^{V^m}$.

(B) If it is in Setting 2, then $\succeq_1^{V_c} = \succeq_2^{V_c}$ and $\succeq_1^{V_{sc}} = \succeq_2^{V_{sc}}$; if X is an ordered vector space, then we also have $\succeq_1^{V_c^m} = \succeq_2^{V_c^m}$ and $\succeq_1^{V_{sc}^m} = \succeq_2^{V_{sc}^m}$.

(C) If it is in Setting 3, then $\succeq_1^{V_c} = \succeq_2^{V_c} = \succeq_3$.

(D) If it is in Setting 4, then $\succeq_1^{V_c} = \succeq_2^{V_c} = \succeq_3 = \succeq_4$.

(E) If it is in Setting 5 and $X = \Re$, then $R_1 = R_2 = R_3 = \succeq_1^V = \succeq_2^V$.

Proof. (A)-(B) Combine Remark 6.1 and Theorem 6.3.

(C) Combine (B) and Theorem A.2.

(D) Combine (C) and Theorem A.5.

(E) Clearly, $R_1 = \succeq_1^V$ in this setting. Now combine (A) and Theorem 2 in Rothschild and Stiglitz [22]. $\blacksquare$

6.2 Partial ordering

It is well-known that $\succeq_1^V$ yields a partial order on $\Delta(O)$ for an outcome space (X, O) in Setting 5 with $X = \Re$. We generalize this result to Setting 1. The generalization is not *verbatim*, but such that it reduces to the classical result in the appropriate setting.

Consider an admissible outcome space (X, O). We say that $\mu, \lambda \in \Delta(O)$ are risk-equivalent, denoted by $\mu \sim \lambda$, if $\mu \circ v^{-1} = \lambda \circ v^{-1}$ for every $v \in V$, i.e.,

the utility distributions $\mu \circ v^{-1}$ and $\lambda \circ v^{-1}$ are identical for every $v \in V$. This is a sensible notion because risk-equivalent lotteries are identical for every risk averse decision-maker. The quotient set $\Delta(O)/\sim$ consists of equivalence classes $[\mu] = \{\mu' \in \Delta(O) \mid \mu' \sim \mu\}$ for $\mu \in \Delta(O)$. Clearly, for $\mu, \lambda \in \Delta(O)$, $\mu \succeq_1^V \lambda$ if and only if $\mu' \succeq_1^V \lambda'$ for all $\mu' \in [\mu]$ and $\lambda' \in [\lambda]$. Consequently, for $\mu, \lambda \in \Delta(O)$, we say $[\mu] \succeq [\lambda]$ if $\mu \succeq_1^V \lambda$.

Theorem 6.5 *If (X, O) is an outcome space in Setting 1, then $\succeq$ is a partial order on $\Delta(O)/\sim$.*

Consider an outcome space (X, O) in Setting 5 with $X = \Re$. Then, the identity function $I \in V$; this holds only when $X = \Re$. If $\lambda \in [\mu]$, then $\mu \sim \lambda$. Consequently, $\mu = \mu \circ I^{-1} = \lambda \circ I^{-1} = \lambda$. So, every equivalence class in $\Delta(O)/\sim$ is a singleton set. Thus, in the Rothschild-Stiglitz setting, $\succeq$ reduces to $\succeq_1^V$ and Theorem 6.5 reduces to the classical result.

Consider an outcome space (X, O) in Setting 1, where O is a path-space. Using Definition 3.5 and Theorem 6.5, it follows that $\succeq$ *also induces a partial order on the quotient set consisting of equivalence classes of processes admissible with respect to $(O, \mathcal{O})$, where processes are declared to be equivalent if their distributions are risk-equivalent.*

6.3 Mean-preservation

It was noted in Footnote 4 that $\mu R_1 \lambda$ implies that μ and λ have the same real mean. This property can be generalized very substantially to a relation representing DTC in Setting 3.

Consider an outcome space (X, O) in Setting 3. The problems in doing the generalization are: (a) to define the mean of $\mu \in \Delta(O)$, i.e., to give meaning to the vector integral $m_\mu := \int_O \mu(dx)\, x$, (b) show that m_μ exists, and (c) show that it is unique.

Let X^* be the set of continuous linear functionals on X. We say that $m_\mu \in X$ is a *Pettis mean* of $\mu \in \Delta(O)$ if $h(m_\mu) = \int_O \mu(dx)\, h(x)$ for every $h \in X^*$. The vector m_μ is the Pettis integral (Pettis [20]) of the identity function on O, which may also be denoted by $\int_O \mu(dx)\, x$.

For instance, if $X = \Re^n$, then the projections $\{\pi_1, \ldots, \pi_n\}$ are a basis for X^* and m_μ is the familiar coordinate-wise mean; if $n = 1$, then m_μ coincides with the familiar (strong) mean $\int_O \mu(dx)\, x$.

The following is a generalized mean-preservation property.

Theorem 6.6 *If (X, O) is an outcome space in Setting 3, then*
  (A) every lottery $\mu \in \Delta(O)$ has a unique Pettis mean $m_\mu \in O$, and
  (B) if $\mu, \lambda \in \Delta(O)$ and $\mu \succeq_1^{V_c} \lambda$, then $m_\mu = m_\lambda$.

This result has implications for risks embodied in processes.

Consider an outcome space (X, O) in Setting 3, where $(O, \mathcal{O})$ is a path-space with $O \subset X \subset \Re^T$. Let $\mathcal{Y} = (\Omega, \mathcal{F}, P; T, y)$ be an admissible process with respect to $(O, \mathcal{O})$. Then, $\hat{y}(\Omega) \subset O$, $\hat{y} : \Omega \to O$ is measurable and $P \circ \hat{y}^{-1} \in \Delta(O)$ is the distribution of $\mathcal{Y}$. The *mean of $\mathcal{Y}$* is defined to be the Pettis mean $m_{P \circ \hat{y}^{-1}}$.[6]

Suppose $\mathcal{Y}_i = (\Omega_i, \mathcal{F}_i, P_i; T, y_i)$ is an admissible process with respect to $(O, \mathcal{O})$ for $i \in \{1, 2\}$. Using Theorem 6.6, $\mathcal{Y}_i$ has a unique mean $m_{P_i \circ \hat{y}_i^{-1}} \in O$ for $i \in \{1, 2\}$. If $\mathcal{Y}_1$ is riskier than $\mathcal{Y}_2$, by which we mean $P_1 \circ \hat{y}_1^{-1} \succeq_1^{V_c} P_2 \circ \hat{y}_2^{-1}$, then $m_{P_1 \circ \hat{y}_1^{-1}} = m_{P_2 \circ \hat{y}_2^{-1}}$. *We conclude that processes with comparable riskiness have identical means.*

6.4 Risk premia

We now relate the riskiness of lotteries to their risk premia in Setting 3. By Theorem 6.6, every $\mu \in \Delta(O)$ has a Pettis mean $m_\mu \in O$. Given $v \in V$, the set of risk premia corresponding to lottery $\mu \in \Delta(O)$ is $\pi(v, \mu) = \{\pi \in X \mid m_\mu - \pi \in O \ \wedge \ v(m_\mu - \pi) = \int_O \mu(dx)\, v(x)\}$.

Given an ordered vector space $(X, \geq)$, we say that $A \geq^* B$ for nonempty sets $A, B \in 2^X$ if $\neg a < b$ for all $a \in A$ and $b \in B$.

Theorem 6.7 *Suppose (X, O) is an outcome space in Setting 3 and $(X, \geq)$ is an ordered vector space. If $\mu, \lambda \in \Delta(O)$ and $\mu \succeq_1^{V_c} \lambda$, then $\pi(v, \mu) \geq^* \pi(v, \lambda)$ for every strictly increasing function $v \in V_c$.*

Specializing this result by setting $X = \Re$, we have the familiar result that $\pi(v, \mu)$ and $\pi(v, \lambda)$ are singleton real numbers and $\pi(v, \mu) \geq \pi(v, \lambda)$.

Consider an outcome space (X, O) in Setting 3, where $(X, \geq)$ is an ordered vector space and $(O, \mathcal{O})$ is a path-space with $O \subset X \subset \Re^T$. Let $\mathcal{Y} = (\Omega, \mathcal{F}, P; T, y)$ be an admissible process with respect to $(O, \mathcal{O})$. Given $v \in V$, the set of risk premia corresponding to $\mathcal{Y}$ is $\pi^*(v, \mathcal{Y}) := \pi(v, P \circ \hat{y}^{-1})$.

If $\mathcal{Y}_i = (\Omega_i, \mathcal{F}_i, P_i; T, y_i)$ *is an admissible process with respect to $(O, \mathcal{O})$ for $i \in \{1, 2\}$ and $\mathcal{Y}_1$ is riskier than $\mathcal{Y}_2$, i.e., $P_1 \circ \hat{y}_1^{-1} \succeq_1^{V_c} P_2 \circ \hat{y}_2^{-1}$, then $\pi^*(v, \mathcal{Y}_1) \geq^* \pi^*(v, \mathcal{Y}_2)$ for every strictly increasing function $v \in V_c$.*

6.5 Aggregation

The theory developed above is sufficient for comparing the riskiness of vector outcome lotteries. But, is it necessary?

Consider an admissible setting (X, O). One might attempt to finesse the problem of comparing the riskiness of lotteries in $\Delta(O)$ as follows. First, convert vector outcomes into real outcomes using a measurable aggregator

---

[6]The Pettis mean may have a pointwise representation in this case. For $t \in T$, if the evaluation $e(t, .) : X \to \Re$ is continuous, then $e(t, .) \in X^*$, and consequently, $m_{P \circ \hat{y}^{-1}}(t) = e(t, m_{P \circ \hat{y}^{-1}}) = \int_O P \circ \hat{y}^{-1}(dx)\, e(t, x) = \int_\Omega P(d\omega)\, e(t, \hat{y}(\omega)) = \int_\Omega P(d\omega)\, y(t, \omega)$.

$a : O \to \Re$. Given an interval $B \supset a(O)$, aggregation yields the distributions $\{\mu \circ a^{-1} \in \Delta(B) \mid \mu \in \Delta(O)\}$. Second, for $\mu, \lambda \in \Delta(O)$, assert that μ is riskier than λ if and only if $\mu \circ a^{-1} R_1 \lambda \circ a^{-1}$, where R_1 is the Rothschild-Stiglitz version of DTC defined over $\Delta(B)$.

Treating the assertion as a postulate merely bypasses the original problem by *fiat*. It is a valid proposition if and only if $\mu \succeq_1^V \lambda$ is equivalent to $\mu \circ a^{-1} R_1 \lambda \circ a^{-1}$ for all $\mu, \lambda \in \Delta(O)$. Given any aggregator satisfying a very mild condition, this equivalence does not hold.

Theorem 6.8 *Consider an outcome space (X, O) in Setting 3. If $a : O \to \Re$ is measurable and not injective, then there exist $\mu, \lambda \in \Delta(O)$ such that $\mu \circ a^{-1} R_1 \lambda \circ a^{-1}$, $\lambda \circ a^{-1} R_1 \mu \circ a^{-1}$, $\neg \mu \succeq_1^{V_c} \lambda$ and $\neg \lambda \succeq_1^{V_c} \mu$.*

As $V_c \subset V$, $\neg \mu \succeq_1^{V_c} \lambda$ and $\neg \lambda \succeq_1^{V_c} \mu$ also imply $\neg \mu \succeq_1^V \lambda$ and $\neg \lambda \succeq_1^V \mu$. As aggregation *per se* involves compression and loss of information, the aggregator being non-injective is an unexceptionable and mild restriction.

Not only does aggregation necessarily yield an incorrect comparison of vector outcome riskiness, it also eliminates interesting substantive aspects of the problem.

For instance, in the problem studied in Section 7.6, aggregation of outcomes would mean resorting to a representative agent and a composite good, thereby suppressing substantive features such as the nature of an agent's preference over allocation vectors and the nature of the planner's preference over profiles of individual utilities and allocations.

Another example is a model where a stream of returns is replaced by a single number, possibly the stream's present value. This compression of outcomes would rob the model of various features of the returns process that could be relevant for preferences, such as volatility, the maximum and minimum processes and various passage times.

7 Applications

We demonstrate in Sections 7.2-7.7 that the theory developed in Section 6 can be used to study a variety of economic applications involving risk embodied in processes. The applications use Theorems 7.1 and 7.2 to predict how a decision-maker's behavior varies with the riskiness of vector outcomes. Theorem 7.1 generalizes an analogous result in Diamond and Stiglitz [8].

7.1 Comparative statics

Consider an admissible outcome space (X, O), $v \in V$ with $v_0 < v_1$, and a family of lotteries $\{\mu(r) \in \Delta(O) \mid r \in \Re\}$ where riskiness increases with $r \in [0, 1]$, i.e., $r_1, r_2 \in [0, 1]$ and $r_2 > r_1$ implies $\mu(r_2) \succeq_2^V \mu(r_1)$.

Let $H : \Re^2 \to \Re$ and $h : \Re^2 \to \Re_+$ satisfy $H(x, r) = F(x, \mu(r) \circ v^{-1}) = \int_{(-\infty, x]} dy\, h(y, r)$ for all $(x, r) \in \Re^2$. Clearly, $H(v_0, r) = 0$ and $H(v_1, r) = 1$ for every $r \in \Re$.

Consider $u : \Re^2 \to \Re$ and $C = [a, b] \subset \Re$ with $a < b$. C is the decision-maker's choice space. If $x \in O$ is the random outcome and the decision-maker chooses $c \in C$, then the resulting utility is $u(v(x), c)$. Define $U : \Re^2 \to \Re$ by $U(r, c) = \int_O \mu(r)(dx)\, u(v(x), c)$. Clearly, $U(r, c) = \int_{[v_0, v_1]} \mu(r) \circ v^{-1}(dx)\, u(x, c) = \int_{[v_0, v_1]} u(x, c)\, dH(x, r) = \int_{[v_0, v_1]} dx\, h(x, r) u(x, c)$. If $\mu(r)$ is the lottery generating the random outcome and the decision-maker chooses $c \in C$, then $U(r, c)$ is the expected utility.

Theorem 7.1 *Suppose (X, O), v, $\{\mu(r) \in \Delta(O) \mid r \in \Re\}$, H, h, C, u and U are as defined above. If*

(a) H is $\mathcal{C}^1$ and h is $\mathcal{C}^2$,

(b) u is $\mathcal{C}^2$; $D_{22}u < 0$ on $[v_0, v_1] \times (a, b)$; $D_{21}u(., c)$ is $\mathcal{C}^1$ for every $c \in C$; and $D_{21}u(v_1, .) = 0$, and

(c) $c : [0, 1] \to (a, b)$ satisfies $U(r, c(r)) \geq U(r, c)$ for every $(r, c) \in [0, 1] \times C$,

then c is differentiable and $D_{211}u \geq 0$ (resp. ≤ 0) implies $Dc \geq 0$ (resp. ≤ 0).

The following is a specialized version of Theorem 7.1.

Theorem 7.2 *Consider the set-up of Theorem 7.1 with $u(x, c) = cf(x) - e(c)$. If (b) is replaced by*

(b') f is $\mathcal{C}^2$ and e is twice differentiable, with $D^2 f \leq 0$, $Df(v_1) = 0$ and $D^2 e > 0$,

then $Dc(r) \leq 0$.

In order to avoid repetition of the above formalism, we assert here that *all the applications in Sections 7.2-7.7 will satisfy the hypotheses of Theorem 7.1, and in some cases, Theorem 7.2.* Most hypothesized properties will simply be postulated, but others will be derived from the specific structure of each problem.

7.2 Auctions with valuation risk

We predict the behavior of bidder 1 (henceforth, "the buyer") in first-price and second-price sealed bid auctions when the buyer's valuation of the prize is determined by a risky vector outcome. For concreteness, let the prize be a natural resource concession and let the risky outcomes refer to the sample paths of a process of returns from the concession; by Theorem 6.8, such vector risks cannot be legitimately reduced to aggregate real risks.

Consider an admissible outcome space (X, O) where $(O, \mathcal{O})$ is an admissible path-space. For the parameter $r \in [0, 1]$, let the prize yield a random

process $\mathcal{Y}(r) = (\Omega, \mathcal{F}, P(r); T, y)$ that is admissible with respect to $(O, \mathcal{O})$. Let $\mu(r) \in \Delta(O)$ be the distribution generated by $\mathcal{Y}(r)$.

Given $x \in O$, let $f \circ v(x)$ be the buyer's gross payoff, where $v \in V$ and f satisfies the conditions listed in Theorem 7.2.[7] Given $r \in [0,1]$, the buyer's expected gross payoff, conditional on winning the prize, is $\int_O \mu(r)(dx) \, f \circ v(x)$. If the buyer's probability of winning the prize is $c \in C := [0,1]$, then the buyer's expected gross payoff is $c \int_O \mu(r)(dx) \, f \circ v(x)$. Let $e(c)$ be the minimum expected cost of bidding to ensure a minimum winning probability c. The buyer's expected net payoff from choosing c is $\int_O \mu(r)(dx) \, u(v(x), c)$, where $u(v, c) = cf(v) - e(c)$. Given $r \in [0,1]$, let $c(r) \in C$ maximize $\int_O \mu(r)(dx) \, u(v(x), .)$. *If e satisfies the conditions of Theorem 7.2, then $Dc(r) \leq 0$, i.e., greater riskiness of the outcome induces the buyer to choose a lower winning probability.* It only remains to derive e and show that it satisfies the conditions of Theorem 7.2.

Given bids $(y_2, \ldots, y_n)$ by the other bidders, let $y^* = \max\{y_2, \ldots, y_n\}$. The buyer wins the prize by bidding b if and only if $y^* \leq b$; for simplicity, we break ties in favor of the buyer. If the buyer wins, then he pays b in a first-price auction and y^* in a second-price auction; otherwise, the buyer pays nothing.

Let G be the distribution function of y^* with $0 \leq \beta_0 < \beta_1 < \infty$, where $\beta_0 = \sup G^{-1}(\{0\})$ and $\beta_1 = \inf G^{-1}(\{1\})$; naturally, G is derived from the buyer's belief about the bidding behavior of the other bidders. If the buyer bids $b \in \Re$, then he wins with probability $G(b)$. Define $b : C \to \Re$ by

$$b(c) = \begin{cases} 0, & \text{if } c = 0 \\ \inf G^{-1}([c, 1]), & \text{if } c \in (0, 1] \end{cases}$$

Consider $c \in (0, 1]$. $G^{-1}([c, 1])$ is bounded below by 0. As G is right-continuous, $G^{-1}([c, 1])$ is closed in $\Re$. So, $b(c) \in G^{-1}([c, 1])$, i.e., $G(b(c)) \geq c$. If $b < b(c)$, then $G(b) < c$. Thus, $b(c)$ is the lowest bid by the buyer that implements c, i.e., yields a winning probability greater than or equal to c.

Lemma 7.3 *The function b is nonnegative, bounded, increasing, measurable and left-continuous.*

There is a simple duality between G and b.

Lemma 7.4 *The function b (resp. G) is continuous if and only if G (resp. b) is strictly increasing on $[\beta_0, \beta_1]$ (resp. C).*

If the buyer bids $b(c)$, then the implied expected cost is

$$g(c) = \begin{cases} b(c)G(b(c)), & \text{for the first-price auction} \\ \int_{[0, b(c)]} y \, dG(y), & \text{for the second-price auction} \end{cases}$$

---

[7]Examples of v include $v(x) = \inf x(T)$ and $v(x) = \int_T dt \, e^{-\delta t} \psi \circ x(t)$ where $\psi : \Re \to \Re$ is concave. The former represents a buyer who is concerned with downside risk while the latter represents a buyer with impatience parameter δ.

Using Lemma 7.3, $g : C \to \Re$ is nonnegative, bounded and increasing. Therefore, it is integrable. In general, g is neither continuous, nor strictly increasing, nor convex. For instance, let $G = 1_{[1,\infty)}$ on $\Re$. Then, $b = 1_{(0,1]}$ on C. Consequently, $g = 1_{(0,1]}$ on C for both auction forms, which is discontinuous and not strictly increasing. As $g(1/2) = 1 > 1/2 = g(1)/2 + g(0)/2$, g is not convex. However, g is more regular in some circumstances.

Lemma 7.5 *If G is strictly increasing on $[\beta_0, \beta_1]$ and continuous, then g is continuous for the first-price and the second-price auctions.*

While $g(c)$ is the least expected cost of implementing c with a deterministic bid, namely $b(c)$, the buyer may reduce the expected cost by randomly implementing c. This amounts to picking a winning probability by using a lottery over C with mean c. The randomization has no effect on the buyer's gross payoff as it is linear in c. We now derive the random implementation of c that minimizes the expected cost of doing so and show that the resulting cost function $e : C \to \Re$ satisfies the conditions of Theorem 7.2.

Suppose the buyer chooses a lottery $\lambda \in \Delta(C)$. The resulting expected winning probability is $m_\lambda := \int_C \lambda(dx)\, x$ and the expected cost is $L(\lambda) := \int_C \lambda(dx)\, g(x)$. As g is nonnegative and bounded, so is L. The set of random implementations of c is $\Delta(C, c) = \{\lambda \in \Delta(C) \mid m_\lambda = c\}$, while $e(c) := \inf\{L(\lambda) \mid \lambda \in \Delta(C, c)\}$ is the minimum expected cost of implementing c. Deriving e is a straightforward matter in special circumstances.

Suppose g is convex. Consider $c \in C$ and $\lambda \in \Delta(C, c)$. As g is convex, $g(c) = g(m_\lambda) \leq \int_C \lambda(dx)\, g(x) = L(\lambda)$. So, $g(c) \leq e(c)$. Since $\delta_c \in \Delta(C, c)$, we have $e(c) \leq L(\delta_c) = g(c)$. Thus, $e = g$.

Suppose g is continuous. As C is a compact metric space, so is $\Delta(C)$ when given the weak* topology (Parthasarathy [18], Theorem II.6.4); see Footnote 10 for more on this topology. By the definition of this topology, L is continuous. If $\Delta(C, c)$ is compact, then there exists $\lambda \in \Delta(C, c)$ such that $e(c) = L(\lambda)$. To show that $\Delta(C, c)$ is compact, it suffices to show that it is a closed subset of $\Delta(C)$. Let $\lambda \in \Delta(C)$ be an accumulation point of $\Delta(C, c)$. Then, there is a sequence $(\lambda_n) \subset \Delta(C, c)$ converging to λ. As the identity mapping on C is continuous, $m_\lambda = \lim_n m_{\lambda_n} = c$. So, $\lambda \in \Delta(C, c)$ and $\Delta(C, c)$ is a closed subset of $\Delta(C)$, as required.

However, as g may be non-convex and discontinuous, the above arguments do not suffice. A general approach is required to derive e without relying on supplementary assumptions regarding g.

The general approach replaces $\Delta(C)$ with the larger set $\Delta(C)^*$ consisting of *finitely additive* lotteries on C. This change has to be accompanied by other modifications: the integrals in the definitions of m_λ and $L(\lambda)$ are now interpreted as Dunford-Schwartz integrals for $\lambda \in \Delta(C)^*$ and we set $\Delta(C, c)^* := \{\lambda \in \Delta(C)^* \mid m_\lambda = c\}$ and $e(c) := \inf\{L(\lambda) \mid \lambda \in \Delta(C, c)^*\}$.[8] A

_____

[8]Since $\Delta(C) \subset \Delta(C)^*$, we have $\Delta(C, c) \subset \Delta(C, c)^*$. Dunford-Schwartz integrals (Dun-

simple argument shows that $e(c) = \inf\{L(\lambda) \mid \lambda \in \Delta(C)^* \ \wedge \ m_\lambda \geq c\}$.[9]

Lemma 7.6 *If $e(c) = \inf\{L(\lambda) \mid \lambda \in \Delta(C,c)^*\}$ for $c \in C$, then*
    (A) $e(c) = L(\lambda) \in \Re$ for some $\lambda \in \Delta(C,c)^*$, and
    (B) $e : C \to \Re$ is nonnegative, bounded, increasing, convex, continuous on $(0,1)$ and twice differentiable Lebesgue almost everywhere.

So, e is well-defined and well-behaved as *per* the requirements of Theorem 7.2. Hence, $Dc(.) \leq 0$ Lebesgue almost everywhere on $[0,1]$.

7.3   Size of regulated utility

Consider an electric utility that is required by the regulator to meet the random demand flow over the period $T = [0,1]$. Failure to do so attracts a penalty. The firm chooses the instantaneous production capacity at time 0, which cannot be adjusted thereafter. Its revenue from the sale of electricity and its variable cost are internalized by the regulator *via* a transfer of the net amount *à la* Laffont and Tirole [14]. In addition, the firm receives a sufficiently large lump-sum return for participating in this arrangement. Given this set-up, the firm chooses capacity to minimize the sum of the fixed cost and the expected penalty. *We show that the chosen capacity increases with the riskiness of the demand process.*

We begin with the formal set-up of the firm's problem. Let $X = \mathcal{C}(T)$. Given $\gamma > 0$, let O be a convex subset of X with $0 \in O$ and $x(T) \subset [0,\gamma]$ for every $x \in O$. The demand for electricity is generated by the random process $\mathcal{Y}(r) = (O, \mathcal{B}(O), \mu(r); T, e)$, where $e : T \times O \to \Re$ is the evaluation function. So, $e(t,x) = x(t) \in [0,\gamma]$ is the instantaneous demand at time t generated by the continuous sample path x. The distribution induced by $\mathcal{Y}(r)$ on its sample paths is $\mu(r)$, with the interpretation that its riskiness increases with parameter $r \in [0,1]$.

Let $C = [a,b]$ be the set of capacity choices with $a \in (0,\gamma)$. Given a demand path $x \in O$ and capacity $c \in C$, $w(x) = \sup x(T)$ is the largest instantaneous demand over T and $\max\{x(t) - c, 0\}$ is the instantaneous deficit at time t. So, the largest instantaneous deficit over T is $\sup_{t \in T} \max\{x(t) - c, 0\} = \max\{\sup x(T) - c, 0\} = \max\{w(x) - c, 0\}$.

Suppose the firm pays a penalty based on the largest instantaneous deficit over T. Given a demand path $x \in O$ and capacity $c \in C$, the penalty is $P(\max\{w(x) - c, 0\})$, where $P : \Re \to \Re$ is a $\mathcal{C}^3$ function with $P(x) = 0$ for $x \leq 0$ and $DP, D^2P, D^3P \geq 0$. Thus, idle capacity is neither rewarded,

---

ford and Schwartz [10], Chapter III) are defined with respect to finitely additive lotteries.

[9]Clearly, $e(c) \geq \inf\{L(\lambda) \mid \lambda \in \Delta(C)^* \ \wedge \ m_\lambda \geq c\}$. Consider $\lambda \in \Delta(C)^*$ with $m_\lambda > c \geq 0$. As $c/m_\lambda < 1$, $h(x) = xc/m_\lambda$ yields the function $h : C \to C$. It follows that $\lambda \circ h^{-1} \in \Delta(C)^*$ and $\int_C \lambda \circ h^{-1}(dy)\,y = \int_C \lambda(dx)\,h(x) = c$. Thus, $\lambda \circ h^{-1} \in \Delta(C,c)^*$ and $L(\lambda \circ h^{-1}) = \int_C \lambda \circ h^{-1}(dy)\,g(y) = \int_C \lambda(dx)\,g \circ h(x) \leq \int_C \lambda(dx)\,g(x) = L(\lambda)$ as g is increasing and $h(x) \leq x$ for every $x \in C$. So, $e(c) \leq \inf\{L(\lambda) \mid \lambda \in \Delta(C)^* \ \wedge \ m_\lambda \geq c\}$.

nor penalized, and penalties increase in severity with the maximum instantaneous deficit. Let $S : \Re \to \Re$ generate the cost $S(c)$ of setting-up capacity $c \in C$. Suppose S is $\mathcal{C}^3$ and $D^2 S > 0$.

Given $r \in [0,1]$, suppose capacity $c = c(r) \in (a,b)$ minimizes the firm's expected cost $\int_O \mu(r)(dx) \, P(\max\{w(x) - c, 0\}) + S(c)$ over C. *We show that, if $r_1, r_2 \in [0,1]$ and $r_1 > r_2$, then $c(r_1) \geq c(r_2)$.*

We first justify referring to $\mathcal{Y}(r)$ as a process. Since X and O do not have measurability structure, we first endow X with the compact-open topology. As T is compact, this topology coincides with the topology of uniform convergence (Dugundji [9], Theorem XII.8.2). By Theorem XII.2.4 in Dugundji [9], e is continuous. Consequently, $e(t, .)$ is continuous, and therefore Borel measurable, for every $t \in T$. So, $\mathcal{Y}(r)$ is indeed a process.

It is easily verified that $w : O \to \Re$ is convex and bounded. For every $\alpha \in \Re$, $w^{-1}((\alpha, \infty)) = \cup_{t \in T}\{x \in O \mid x(t) > \alpha\} = \cup_{t \in T}\{x \in O \mid e(x,t) > \alpha\} = \cup_{t \in T} e(., t)^{-1}((\alpha, \infty))$ is open in O as e is continuous. Thus, w is lower semicontinuous. Therefore, w is Borel measurable.

So, $v := -w \in V$, $v(O) \subset [-\gamma, 0]$ and $v_1 = \sup v(O) = -\inf w(O) = 0$ as sample path $0 \in O$. Define $g : \Re \to \Re$ by $g(x) = \min\{x, 0\}$. Then, for $r \in [0,1]$, $c = c(r) \in (a,b)$ maximizes

$$U(r,c) = \int_O \mu(r)(dx) \, u(v(x), c)$$

over C, where $u : \Re^2 \to \Re$ is given by $u(v,c) = -S(c) - P \circ (I - g)(-v - c)$ and I is the identity mapping. Although this problem resembles the problem of Theorem 7.1, we cannot apply this theorem directly as u is not sufficiently regular. Our indirect method for analyzing $c(.)$ proceeds in three steps.

1. We uniformly approximate g by a sequence (g_n) of smooth functions. Each g_n generates a uniform approximation u_n of u over the set $v(O) \times C$. Each u_n generates an approximate version of the firm's problem and a solution $c(.; n)$ of this approximate problem.

2. We use Theorem 7.1 to show that $c(r_1; n) \geq c(r_2; n)$ for all $n \in \mathcal{N}$ and $r_1, r_2 \in [0,1]$ with $r_1 > r_2$.

3. We generate cluster points $c^*(r_1)$ and $c^*(r_2)$ of the sequences $(c(r_1; n))$ and $(c(r_2; n))$ respectively, and show that $c(r_1) = c^*(r_1) \geq c^*(r_2) = c(r_2)$, as required.

The following approximation is the key to the first step. It is derived by smoothly pasting smooth functions over various parts of the domain.

Lemma 7.7 *Consider $b \in \Re$ and $\epsilon > 0$. If $u : \Re \to \Re$ is given by $u(x) = \min\{x - b, 0\}$, then there exists a concave, increasing $\mathcal{C}^\infty$ function $h : \Re \to \Re$ such that $\sup\{|h(x) - u(x)| \mid x \in \Re\} \leq \epsilon$, $h \geq u$, $Dh = 1$ on $(-\infty, b - 2\epsilon]$, $Dh \in (0,1)$ on $(b - 2\epsilon, b)$ and $h = Dh = 0$ on $[b, \infty)$.*

Fix $N \in \mathcal{N}$ such that $2/N < a$ and consider $n \geq N$. Lemma 7.7 implies that there exists a concave, increasing $\mathcal{C}^\infty$ function $g_n : \Re \to \Re$ such that $\sup\{|g_n(x) - g(x)| \mid x \in \Re\} \leq 1/n$, $g_n \geq g$, $Dg_n = 1$ on $(-\infty, -2/n]$, $Dg_n \in (0,1)$ on $(-2/n, 0)$ and $g_n = Dg_n = 0$ on $\Re_+$. Set $g_n = g_N$ for $n < N$. Given $n \in \mathcal{N}$, define $u_n : \Re^2 \to \Re$ by $u_n(v,c) = -S(c) - P \circ (I - g_n)(-v - c)$. Clearly, u_n is $\mathcal{C}^3$. The following result completes the first step.

Lemma 7.8 *The sequence (u_n) converges uniformly to u on $v(O) \times C$.*

Let $c(r; n) \in (a, b)$ maximize $U(r, .; n) := \int_O \mu(r)(dx)\, u_n(v(x), .)$ over C for $r \in [0, 1]$ and $n \in \mathcal{N}$. The next result completes the second step.

Lemma 7.9 *If $n \geq N$, $r_1, r_2 \in \Re$ and $r_1 > r_2$, then $c(r_1; n) \geq c(r_2; n)$.*

As C is compact, the Bolzano-Weierstrass property implies the existence of subsequences $(c(r_1, j)) \subset (c(r_1, n))$ and $(c(r_2, j)) \subset (c(r_2, n))$ converging to $c^*(r_1) \in C$ and $c^*(r_2) \in C$ respectively. By the above arguments, $c(r_1; j) \geq c(r_2; j)$ for all but a finite set of j. Consequently, $c^*(r_1) \geq c^*(r_2)$.

Lemma 7.10 *For $r \in \{r_1, r_2\}$, $c^*(r)$ maximizes $U(r, .)$ over C.*

As $U(r, .)$ is strictly concave, its maximum over C is unique. Therefore, Lemma 7.10 implies $c(r_1) = c^*(r_1) \geq c^*(r_2) = c(r_2)$, as claimed.

7.4   Optimal inventory management

The previous section featured the problem of managing the production of a non-durable commodity. The same formalism can be applied to the problem of managing the inventory of a durable commodity. Since the formal analysis is identical, we shall only re-interpret the formalism of Section 7.3.

Consider a firm that produces a durable commodity. Let ct, with $c > 0$, be the firm's output over the period $[0, t]$. Let $D(t)$ be the total orders that arrive over the period $[0, t]$. So, the stock-out at time t is $\max\{D(t) - ct, 0\}$. In order to discourage stock-outs, the firm imposes a penalty on the manager, which depends on the largest time-average of stock-outs over the period $T = [0, 1]$. Thus, the penalty is based on $\sup_{t \in T} \max\{D(t)/t - c, 0\}$.

The manager's instrument for controlling stock-outs is the production parameter c. The manager has to pay cost $S(c)$ for choosing c.

Let T, X, O, w, C, P, S and $\mathcal{Y}(r)$ be as specified in Section 7.3. The only change is that $\mathcal{Y}(r)$ now represents the firm's time-average of orders flow, i.e., $e(x, t) = x(t) = D(t)/t$ for $t \in T$ and $x \in O$.

Given $r \in [0, 1]$, suppose the production technology $c = c(r) \in (a, b)$ minimizes the manager's expected penalty $\int_O \mu(r)(dx)\, P(\max\{w(x) - c, 0\}) + S(c)$ over C. As shown in Section 7.3, *if $r_1, r_2 \in [0, 1]$ and $r_1 > r_2$, then $c(r_1) \geq c(r_2)$, i.e., greater riskiness of the time-average of orders process induces the manager to choose a higher production rate.*

7.5 Portfolio choice

Consider an investor with wealth 0 who invests in a risky asset with market value 1. A share $c \in C := [0, 1]$ of the asset costs c and yields a share c of the random dividends from it that arrive over the period $T = [0, 1]$. Let the sample paths of the dividend process belong to $O \subset X := \mathcal{C}(T)$. A sample path $x \in O$ yields dividend $x(t)$ at time $t \in T$. Let $v \in V$ be the statistic used by the investor to evaluate sample paths. We assume that v is homogeneous of degree 1. Examples of v include $v(x) = x(\tau)$ for some $\tau \in T$, $v(x) = \inf x(T)$ and $v(x) = \int_T dt \, e^{-\delta t} x(t)$.

Given $c \in C$ and $x \in O$, the investor receives the dividend path cx. The resulting gross payoff, $f(v(cx) - cv_1) = f(c(v(x) - v_1))$, depends on the variation from the reference statistical evaluation v_1. Let $f : \Re \to \Re$ be $\mathcal{C}^3$ with $Df > 0$, $D^2 f < 0$ and $D^3 f > 0$ on $[v_0 - v_1, 0)$ and $Df(0) = 0$.

The investor's purchase of a share c is financed by selling $Q(c)$ bonds at price $c/Q(c)$. Each bond yields a gross return $\alpha > 0$ at time 1.

Given $0 < \beta < \gamma$, let $G : [\beta, \gamma] \to \Re$ be the demand function for the investor's bonds; if the bond price is p, then $G(p)$ is the number of bonds sold. Suppose G is $\mathcal{C}^2$, with $G(\beta) = 1/\beta$, $G(\gamma) = 0$, $DG < 0$, $D^2 G < 0$ and $G(\beta) + \beta DG(\beta) < 0$. Then, there is a unique $\mathcal{C}^2$ implicit function $P : C \to [\beta, \gamma]$ such that $c = G(P(c))P(c)$ for every $c \in C$. This implies $P(0) = \gamma$, $P(1) = \beta$, $P > 0$, $DP < 0$ and $D^2 P < 0$. Setting $Q = G \circ P$, it follows that Q is $\mathcal{C}^2$, with $Q \geq 0$, $DQ > 0$ and $D^2 Q > 0$. *So, a share c of the risky asset is financed by selling $Q(c)$ bonds at price $P(c)$.*

The investor's net payoff is generated by the function $u : v(O) \times C \to \Re$ given by $u(v, c) = f(c(v - v_1)) - \alpha Q(c)$. The investor chooses $c \in C$ to maximize $\int_O \mu(r)(dx) \, u(v(x), c) = \int_O \mu(r)(dx) \, f(c(v(x) - v_1)) - \alpha Q(c)$.

As $D_{211} u(v, c) = c[c(v - v_1)D^3 f(c(v - v_1)) + 2D^2 f(c(v - v_1))] < 0$, $D_{22} u(v, c) = (v - v_1)^2 D^2 f(c(v - v_1)) - \alpha D^2 Q(c) < 0$ and $D_{21} u(v_1, c) = Df(0) = 0$, Theorem 7.1 implies that $Dc(r) \leq 0$, i.e., greater riskiness of the asset implies a smaller investment in it by the investor.

7.6 Allocation of a public good

We study a planner's problem of using a given resource to provide a period-specific public good in each of two periods, say 1 and 2. More precisely, the two goods are excludable across periods but each period-specific good is non-excludable and non-rival within the relevant period.

Let I be a finite set of agents. An agent's allocation in a period consists of the period-specific public good provided by the planner and a vector of other period-specific goods. The planner's resource constraint is modeled by interpreting $\phi(c) \in \Re_+$ as the maximum amount of the public good that the planner can provide in period 1 if $c \in C \subset \Re_+$ is the amount of the public good to be supplied in period 2. The allocation of other goods in period 1

is determined by a known outcome $y \in O$ and the corresponding allocation in period 2 is determined by a randomly generated outcome $x \in O$.

Given outcomes $x, y \in O$ and the planner's decision $c \in C$, the aggregate social welfare is $w(v(y), \phi(c)) + w(v(x), c)$, where $v : O \to \Re$ is the function used by the planner to evaluate outcomes in O. Since y is given, we suppress it and set $\psi(c) = -w(v(y), \phi(c))$. Therefore, the social welfare function $u : v(O) \times C \to \Re$ is given by $u(v, c) = w(v, c) - \psi(c)$. If the period 2 outcome is $x \in O$ and the planner's decision is $c \in C$, then social welfare is $u(v(x), c) = w(v(x), c) - \psi(c)$.

We consider alternative specifications of X, O, x and v and analyze how the optimal c varies with the riskiness of x.

Example 7.11 *Let $X = (\Re^l)^I$ and $O \subset (\Re_+^l)^I$. The random outcome is $x = (x_i)_{i \in I} \in O$, where x_i is agent i's allocation of private goods in period 2. Let agent i's valuation of x_i be $f_i(x_i)$, where $f_i : \Re_+^l \to \Re$ is continuous and concave. Define $v : O \to \Re$ by $v(x) = \min\{f_i(x_i) \mid i \in I\}$ for $x = (x_i)_{i \in I}$.*

In the next example, instead of directly facing a risky private goods allocation $x_i \in \Re_+^l$, agent i indirectly faces this risk *via* a risky period 2 aggregate endowment $x \in \Re_{++}^l$ that the planner allocates optimally among the agents using the Rawlsian criterion.

Example 7.12 *Suppose $X = \Re^l$, $O \subset \Re_{++}^l$ and $x \in O$ is the aggregate endowment of private goods in period 2. The planner allocates x among the agents. Given x, the set of feasible allocations is $S(x) = \{(x_i)_{i \in I} \in (\Re_+^l)^I \mid \sum_{i \in I} x_i = x\}$. Given f_i as in Example 7.11, define $v : O \to \Re$ by $v(x) = \sup\{\min\{f_i(x_i) \mid i \in I\} \mid (x_i)_{i \in I} \in S(x)\}$.*

The next example assumes that an agent's allocation in a given period consists of the planner's public good and a vector of other public goods; private goods are suppressed only for notational simplicity.

Example 7.13 *Suppose $X = \Re^l$, $O \subset \Re_+^l$ and $x \in O$ is a vector of other public goods provided in period 2. Let $f_i(x)$ be agent i's valuation of x. Given f_i as in Example 7.11, define $v : O \to \Re$ by $v(x) = \min\{f_i(x) \mid i \in I\}$.*

The next result allows the application of Theorem 7.1 to our examples.

Lemma 7.14 *In Examples 7.11-7.13, if O is compact, then $v \in V$.*

Suppose O is compact. Then, $v \in V$ by Lemma 7.14. Suppose w and ϕ are $\mathcal{C}^2$, $D_2 w > 0$, $D^2 \phi < 0$, $D_{22} w < 0$, $D_1 w(v_1, .) = 0$ and $D_{21} w(., c)$ is $\mathcal{C}^1$ for every $c \in C$. Then, u satisfies hypothesis (b) of Theorem 7.1.

If $D_{211} w \geq 0$, then Theorem 7.1 implies that, in all the above examples, the planner's optimal response to an increase in riskiness of the random outcome $x \in O$ is to increase the second period public good allocation c.

Replacing the Rawlsian criterion with the Benthamite criterion, v in Examples 7.11-7.13 is given by $v(x) = \sum_{i \in I} f_i(x_i)$, $v(x) = \sup\{\sum_{i \in I} f_i(x_i) \mid (x_i)_{i \in I} \in S(x)\}$ and $v(x) = \sum_{i \in I} f_i(x)$ respectively. It is easily confirmed that our result is preserved in these cases.

7.7   Moral hazard in teams

The following example provides a template for numerous moral hazard problems in team settings.

Consider a principal who might suffer a loss $\alpha > 0$, say on account of an accident. The probability of loss is $Q(v(x), c)$, where $c \in C = [a, b]$ is the precaution exercised by the principal, $x = (x_1, \ldots, x_n)$ is the profile of precautions chosen by the n agents in the team and $v(x) = \min\{x_1, \ldots, x_n\}$ is the effective precaution exercised by the agents. Suppose choosing precaution c imposes cost c on the principal. Thus, the principal selects $c \in C$ to maximize $\int_O \mu(r)(dx)\, u(v(x), c)$, where $u(v, c) = -c - \alpha Q(v, c)$ and $\mu(r)$ is the principal's belief about the precautions chosen by the agents.

Suppose $Q : v(O) \times C \to [0, 1]$ is $\mathcal{C}^2$ and decreasing, $D_{22}Q > 0$ on $v(O) \times (a, b)$, $D_{21}Q(., c)$ is $\mathcal{C}^1$ for every $c \in C$ and $D_1 Q(v_1, .) = 0$. Let $X = \Re^n$ and let O be compact. *If $c : [0, 1] \to (a, b)$ is as in Theorem 7.1, then the sign of $Dc(r)$ is the opposite of the sign of $D_{211}Q$.*

8   Concluding remarks

Given *any* path-space, we have shown the equivalence of two criteria for comparing the riskiness of random processes that generate distributions over the given path-space; indeed, our main result yields many versions of this equivalence. This equivalence is a consequence of a more abstract result showing the equivalence of two criteria for comparing the riskiness of lotteries over *any* real vector space.

While one of the equivalent criteria is the decision-theoretic touchstone of the theory of comparative riskiness, the other criterion is very tractable in applications. We substantiate this claim by using the second criterion in six economic applications. We also generalize various supplementary results of the Rothschild-Stiglitz theory to vector outcomes settings.

A   Appendix: Classical results

Consider Setting 3. A Markov kernel on O is a function $P : O \times \mathcal{B}(O) \to \Re$ such that $P(x, .) \in \Delta(O)$ for every $x \in O$ and $P(., E)$ is measurable for every $E \in \mathcal{B}(O)$. A dilatation on O is a Markov kernel P on O such that $f(.) = \int_O P(., dx)\, f(x)$ for every continuous affine function $f : O \to \Re$.

Definition A.1 *In Setting 3, given $\mu, \lambda \in \Delta(O)$, we say that $\mu \succeq_3 \lambda$ if there is a dilatation P on O such that $\mu(.) = \int_O \lambda(dx) \, P(x,.)$.*

The following result is Theorem 2 in Strassen [28].

Theorem A.2 *In Setting 3, $\succeq_1^{V_c} = \succeq_3$.*

In order to state the other equivalence result, we first define a martingale.

Definition A.3 *Let X be a separable Banach space. An ordered pair (f, g) is an X-valued martingale if there is a probability space $(\Omega, \mathcal{F}, Q)$ such that*
(a) $f, g \in L_1(\Omega, \mathcal{F}, Q; X)$
(b) $\sigma(f) := \{f^{-1}(B) \mid B \in \mathcal{B}(X)\} \subset \sigma(g) := \{g^{-1}(B) \mid B \in \mathcal{B}(X)\}$, and
(c) $\int_E Q(dx) \, f(x) = \int_E Q(dx) \, g(x)$ for every $E \in \sigma(f)$.

Condition (a) means that f and g are Bochner integrable random elements with values in X. Condition (b) means that $(\sigma(f), \sigma(g))$ is a filtration. By definition, (f, g) is adapted to this filtration. Condition (c) means that f is (a version of) the expectation of g conditional on f, i.e., $f = E_{\sigma(f)} g$. Setting $h = g - f$, we have $E_{\sigma(f)} h = E_{\sigma(f)} g - f = 0$. Therefore, we have the representation $g = f + h$, where $E_{\sigma(f)} h = 0$, i.e., g is f plus noise. This interpretation motivates the following definition.

Definition A.4 *In Setting 4, given $\mu, \lambda \in \Delta(O)$, we say that $\mu \succeq_4 \lambda$ if there exists an X-valued martingale (f, g) with $Q \circ f^{-1} = \lambda$ and $Q \circ g^{-1} = \mu$.*

The following result is a consequence of Theorem 8 in Strassen [28].

Theorem A.5 *In Setting 4, $\succeq_1^{V_c} = \succeq_4$.*

B Appendix: Proofs

Proof of Theorem 3.4 As $(O, \mathcal{O})$ is admissible, $\mathcal{E} = (O, \mathcal{O}, P; T, e)$ is a process. As $\hat{e}$ is the identity function on O, it is measurable. Therefore, $\mathcal{E}$ is admissible with respect to $(O, \mathcal{O})$ and $P = P \circ \hat{e}^{-1}$. ∎

Proof of Theorem 6.5 As $\succeq_1^V$ is reflexive and transitive, so is $\succeq$. We show that $\succeq$ is antisymmetric. Let $[\mu] \succeq [\lambda]$ and $[\lambda] \succeq [\mu]$ for $\mu, \lambda \in \Delta(O)$. To show that $[\mu] = [\lambda]$, it suffices to show that $\mu \sim \lambda$, i.e., $\mu \circ v^{-1} = \lambda \circ v^{-1}$ for every $v \in V$.

Consider $v \in V$. By the definition of $\succeq$, we have $\mu \succeq_1^V \lambda$ and $\lambda \succeq_1^V \mu$. Using Corollary 6.4, $\mu \succeq_2^V \lambda$ and $\lambda \succeq_2^V \mu$. Therefore, $c(y) := \int_{(-\infty, y]} dx \, [F(x, \mu \circ v^{-1}) - F(x, \lambda \circ v^{-1})] = 0$ for every $y \in \Re$. Consequently, $Dc(y) = 0$ for every $y \in \Re$. If y is a continuity point of $F(., \mu \circ v^{-1}) - F(., \lambda \circ v^{-1})$, then $0 = Dc(y) = F(y, \mu \circ v^{-1}) - F(y, \lambda \circ v^{-1})$ (Bartle [2], Theorem 31.8). Suppose

y is a discontinuity point of $F(., \mu \circ v^{-1}) - F(., \lambda \circ v^{-1})$. The set of such discontinuity points is countable (Billingsley [4], Theorem 10.2). Consequently, there exists a sequence of continuity points $(y_n) \subset (y, \infty)$ converging to y. As $F(y_n, \mu \circ v^{-1}) - F(y_n, \lambda \circ v^{-1}) = 0$ for every n and distribution functions are right-continuous, we have $F(y, \mu \circ v^{-1}) - F(y, \lambda \circ v^{-1}) = 0$. It follows that $F(., \mu \circ v^{-1}) = F(., \lambda \circ v^{-1})$.

$\mathcal{P} = \{(-\infty, y] \mid y \in \Re\}$ is a π-system and $\mathcal{L} = \{E \in \mathcal{B}(\Re) \mid \mu \circ v^{-1}(E) = \lambda \circ v^{-1}(E)\}$ is a λ-system (Billingsley [4], Page 41). By the above argument, $\mathcal{P} \subset \mathcal{L}$. By Dynkin's theorem (Billingsley [4], Theorem 3.2), $\mathcal{B}(\Re) = \sigma(\mathcal{P}) \subset \mathcal{L}$, i.e., $\mu \circ v^{-1} = \lambda \circ v^{-1}$. ∎

Proof of Theorem 6.6 (A) Consider $\mu \in \Delta(O)$. If $a, b \in X$ and $h(a) = \int_O \mu(dx) h(x) = h(b)$ for every $h \in X^*$, then $h(a - b) = h(a) - h(b) = 0$ for every $h \in X^*$. Consequently, $a = b$ (Dunford and Schwartz [10], Corollary V.2.13). So, if m_μ exists, then it is unique. We now show that $m_\mu \in O$ exists.

Define $H : X \to \Re^{X^*}$ by $H(x) = (h(x))_{h \in X^*}$. Give $\Re^{X^*}$ the product topology. H is continuous as all its component functions are continuous. If $H(x) = H(y)$ for $x, y \in X$, then $h(x - y) = h(x) - h(y) = 0$ for every $h \in X^*$, which implies $x = y$. So, H is injective. As O is compact and $\Re^{X^*}$ is Hausdorff, H imbeds O in $\Re^{X^*}$. So, $H(O)$ is closed in $\Re^{X^*}$ and metrizable.

Consider $\mu \in \Delta(O)$ with finite support. The linearity of $h \in X^*$ implies

$$\int_O \mu(dz) h(z) = \sum_{z \in \operatorname{supp}\mu} \mu(\{z\})h(z) = h\left(\sum_{z \in \operatorname{supp}\mu} \mu(\{z\})z\right) \qquad (3)$$

Setting $m_\mu = \sum_{z \in \operatorname{supp}\mu} \mu(\{z\})z$, we have $m_\mu \in O$ as O is convex and $\operatorname{supp}\mu \subset O$. Thus, $H(m_\mu) \in H(O)$ for every $\mu \in \Delta(O)$ with $|\operatorname{supp}\mu| < \infty$.

Consider $\mu \in \Delta(O)$. As O is compact and metric, it is separable. Consequently, there exists a sequence $(\mu_n) \subset \Delta(O)$ converging to μ in the $\mathcal{C}(O)$ topology such that $|\operatorname{supp}\mu_n| < \infty$ for every $n \in \mathcal{N}$ (Parthasarathy [18], Theorem II.6.3).[10] By the above argument, m_{μ_n} exists, $m_{\mu_n} \in O$ and $H(m_{\mu_n}) \in H(O)$ for every $n \in \mathcal{N}$. Using (3) and the definition of convergence in the $\mathcal{C}(O)$ topology,

$$\lim_{n\uparrow\infty} h(m_{\mu_n}) = \lim_{n\uparrow\infty} \int_O \mu_n(dz) h(z) = \int_O \mu(dz) h(z)$$

for every $h \in X^*$. Thus, $\lim_{n\uparrow\infty} H(m_{\mu_n}) = (\int_O \mu(dz) h(z))_{h \in X^*}$. As the sequence $(H(m_{\mu_n})) \subset H(O)$ and $H(O)$ is closed in $\Re^{X^*}$ and metrizable, we have $(\int_O \mu(dz) h(z))_{h \in X^*} \in H(O)$. As H imbeds O in $\Re^{X^*}$, there exists a

---

[10]The $\mathcal{C}(O)$ topology is the weakest topology on $\Delta(O)$ that makes the functional $\mu \mapsto \int_O \mu(dx) f(x)$ continuous for every $f \in \mathcal{C}(O)$. It belongs to the class of weak* topologies in analysis and is usually called "the weak topology" in probability theory.

unique $x \in O$ such that $H(x) = (\int_O \mu(dz)\, h(z))_{h \in X^*}$. By the definition of H, we have $h(x) = \int_O \mu(dz)\, h(z)$ for every $h \in X^*$. Set $m_\mu = x$.

(B) Suppose $\mu \succeq_1^{V_c} \lambda$. Consider $h \in X^*$. It is straightforward to verify that the restrictions of h and $-h$ to O belong to V_c. It follows that $\mu \succeq_1^{V_c} \lambda$ implies $\int_O \mu(dx)\, h(x) = \int_O \lambda(dx)\, h(x)$ for every $h \in X^*$. By (A), there exist unique vectors $m_\mu, m_\lambda \in O$ such that $h(m_\mu) = \int_O \mu(dx)\, h(x)$ and $h(m_\lambda) = \int_O \lambda(dx)\, h(x)$ for every $h \in X^*$. Thus, $h(m_\mu - m_\lambda) = h(m_\mu) - h(m_\lambda) = 0$ for every $h \in X^*$. So, $m_\mu = m_\lambda$ (Dunford and Schwartz [10], Corollary V.2.13).

If $\mu \succeq_1^V \lambda$, then $\mu \succeq_1^{V_c} \lambda$ and the above argument applies. $\blacksquare$

Proof of Theorem 6.7 Suppose $\mu \succeq_1^{V_c} \lambda$ for $\mu, \lambda \in \Delta(O)$ and $v \in V_c$ is strictly increasing. By Theorem 6.6, μ and λ have unique Pettis means $m_\mu, m_\lambda \in O$ and $m_\mu = m_\lambda$.

As O is convex, it is connected. As v is continuous, $v(O) \subset \Re$ is an interval. As v is bounded, $\int_O \mu(dx)\, v(x)$ exists and $\int_O \mu(dx)\, v(x) = v(c)$ for some $c \in O$. As $m_\mu - c \in \pi(v, \mu)$, we have $\pi(v, \mu) \neq \emptyset$. Similarly, $\pi(v, \lambda) \neq \emptyset$.

Consider $a \in \pi(v, \mu)$ and $b \in \pi(v, \lambda)$. Since $\mu \succeq_1^{V_c} \lambda$, $v(m_\mu - a) = \int_O \mu(dx)\, v(x) \leq \int_O \lambda(dx)\, v(x) = v(m_\lambda - b)$. As v is strictly increasing, $\neg m_\mu - a > m_\lambda - b$, and so, $\neg a < b$. Consequently, $\pi(v, \mu) \geq^* \pi(v, \lambda)$. $\blacksquare$

Proof of Theorem 6.8 Consider $x \in X$ and $y \in X - \{x\}$. Then, there is a continuous linear functional $g : X \to \Re$ such that $g(x) > g(y)$ (Dunford and Schwartz [10], Corollary V.2.13). As g is continuous, $U(y) := g^{-1}((-\infty, g(x)))$ is an open neighborhood of y that does not include x. So, $X - \{x\} = \cup_{y \in X - \{x\}} U(y)$ is open in X. Therefore, $\{x\}$ is closed in X for every $x \in X$.

As a is not injective, there exist $x, y \in O$ such that $x \neq y$ and $a(x) = a(y)$. By the above argument, there is a continuous linear functional $f : X \to \Re$ such that $f(x) > f(y)$. The restrictions of f and $-f$ to O belong to V_c. As $\{x\}$ and $\{y\}$ are closed subsets of O, they are Borel sets. So, $\delta_x, \delta_y \in \Delta(O)$. As a is bounded, $B := [\inf a(O), \sup a(O)] \subset \Re$ and $\delta_x \circ a^{-1}, \delta_y \circ a^{-1} \in \Delta(B)$. As $\delta_x \circ a^{-1} = \delta_y \circ a^{-1}$, we have $\delta_x \circ a^{-1} \succeq_1^V \delta_y \circ a^{-1}$ and $\delta_y \circ a^{-1} \succeq_1^V \delta_x \circ a^{-1}$. Since $\int_O \delta_x(dz)\, f(z) > \int_O \delta_y(dz)\, f(z)$ and $\int_O \delta_x(dz)\, (-f)(z) < \int_O \delta_y(dz)\, (-f)(z)$, we have $\neg \delta_x \succeq_1^{V_c} \delta_y$ and $\neg \delta_y \succeq_1^{V_c} \delta_x$. $\blacksquare$

Proof of Theorem 7.1 Define $\phi : \Re \times \Re^2 \to \Re$ and $\Phi : [v_0, v_1] \times \Re \to \Re$ by $\phi(x, (r, c)) = h(x, r)u(x, c)$ and $\Phi(x, r) = \int_{[v_0, x]} dy\, H(y, r)$. Suppose (i) ϕ is $\mathcal{C}^2$, (ii) U is $\mathcal{C}^2$ on $[0, 1] \times C$, and (iii) $\Phi(x, .)$ is differentiable for every $x \in [v_0, v_1]$ and $D_2\Phi \geq 0$ on $[v_0, v_1] \times [0, 1]$. We shall verify these facts later.

Consider $r \in [0, 1]$. Given (ii), (b) and (c), $D_2U(r, c(r)) = 0$ and $D_{22}U(r, c(r)) < 0$. The implicit function theorem implies that c is $\mathcal{C}^1$ in a neighborhood of r. Using (ii), $D_{21}U(r, c(r)) + D_{22}U(r, c(r))Dc(r) = 0$. Consequently, $Dc(r)$ and $D_{21}U(r, c(r))$ have the same sign.

Let $(r, c) \in [0, 1] \times C$. Using (i) and integrating by parts, $D_2U(r, c) =$

$\int_{[v_0,v_1]} dx\, h(x,r) D_2 u(x,c) = D_2 u(v_1,c) - \int_{[v_0,v_1]} dx\, D_{21} u(x,c) H(x,r)$.[11] Using (a) and (b), $D_{21} U(r,c) = -\int_{[v_0,v_1]} dx\, D_{21} u(x,c) D_2 H(x,r)$. Integrating by parts and using (b), $D_{21} U(r,c) = \int_{[v_0,v_1]} dx\, D_{211} u(x,c) D_2 \Phi(x,r)$. Given (iii), $D_{211} u \geq 0$ (resp. ≤ 0) implies $D_{21} U \geq 0$ (resp. ≤ 0), in particular, $D_{21} U(r,c(r)) \geq 0$ (resp. ≤ 0), and consequently, $Dc(r) \geq 0$ (resp. ≤ 0).

We now verify claims (i)-(iii).

(i) Using (a), (b) and Theorem XIII.7.1 in Lang [15], ϕ is $\mathcal{C}^2$.

(ii) As $[v_0,v_1]$ is bounded, $\phi(.,(r,c))$ is bounded on $[v_0,v_1]$. So, $U(r,c) = \int_{[v_0,v_1]} dx\, \phi(x,(r,c))$ exists. Using (i), $DU(r,c) = \int_{[v_0,v_1]} dx\, D_2 \phi(x,(r,c))$ and analogous arguments applied to the component functions of DU imply that $D^2 U(r,c) = \int_{[v_0,v_1]} dx\, D_{22} \phi(x,(r,c))$, i.e., U is twice differentiable. Since $D_{22}\phi$ is continuous on the compact set $[v_0,v_1] \times [0,1] \times C$, the component functions are bounded on this set. Consequently, $D^2 U$ is continuous on $[0,1] \times C$ (Lang [15], Exercise VI.10.13). Therefore, U is $\mathcal{C}^2$ on $[0,1] \times C$.

(iii) Using (a), $D_2 \Phi(x,.) = \int_{[v_0,x]} dy\, D_2 H(y,.)$ for every $x \in [v_0,v_1]$. As $r_1, r_2 \in [0,1]$ and $r_2 > r_1$ implies $\mu(r_2) \succeq_2^V \mu(r_1)$, $\Phi(x,.)$ is increasing on $[0,1]$ for every $x \in [v_0,v_1]$. So, $D_2\Phi \geq 0$ on $[v_0,v_1] \times [0,1]$. ∎

Proof of Theorem 7.2 Specializing the proof of Theorem 7.1, we have $\int_{[v_0,v_1]} dx\, h(x,r) f(x) - De(c(r)) = D_2 U(r,c(r)) = 0$ and $-D^2 e(c(r)) = D_{22} U(r,c(r)) < 0$. As De is differentiable and strictly increasing, it has a differentiable inverse. So, $c(r) = (De)^{-1}(\int_{[v_0,v_1]} dx\, h(x,r) f(x))$. The chain rule and (a) imply that $c(.)$ is differentiable. So, $\int_{[v_0,v_1]} dx\, D_2 h(x,r) f(x) - D^2 e(c(r)) Dc(r) = 0$. Therefore, $Dc(r)$ and $\int_{[v_0,v_1]} dx\, D_2 h(x,r) f(x)$ have the same sign. Twice integrating by parts and using (a) and (b'), we have $\int_{[v_0,v_1]} dx\, D_2 h(x,r) f(x) = \int_{[v_0,v_1]} dx\, D^2 f(x) \int_{[v_0,x]} dy\, D_2 H(y,r) \leq 0$. ∎

Proof of Lemma 7.3 As $\beta_0 \in \Re_+$ and $\beta_1 \in \Re$, b is nonnegative and bounded.

Let $c_1, c_2 \in C$ such that $c_1 < c_2$. If $c_1 = 0$, then $b(c_1) = 0 \leq b(c_2)$. If $c_1 > 0$, then $G^{-1}([c_2,1]) \subset G^{-1}([c_1,1])$; consequently, $b(c_1) = \inf G^{-1}([c_1,1]) \leq \inf G^{-1}([c_2,1]) = b(c_2)$, i.e., b is increasing and therefore measurable.

If $\alpha < 0$, then $b^{-1}((\alpha,\infty)) = C$, which is open in C. Consider $\alpha \geq 0$. Let $c \in b^{-1}((\alpha,\infty))$. Then, $c \in (0,1]$ and $\inf G^{-1}([c,1]) = b(c) > \alpha$. Consequently, $G(\alpha) < c$. Thus, $b^{-1}((\alpha,\infty)) \subset (0,1] \cap (G(\alpha),\infty)$. Conversely, suppose $c \in (0,1] \cap (G(\alpha),\infty)$. As $G(b(c)) \geq c$, we have $G(\alpha) < G(b(c))$. As G is increasing, $\alpha < b(c)$. Thus, $b^{-1}((\alpha,\infty)) \supset (0,1] \cap (G(\alpha),\infty)$. So, $b^{-1}((\alpha,\infty)) = (0,1] \cap (G(\alpha),\infty)$, which is open in C. Therefore, b is left-continuous. ∎

Proof of Lemma 7.4 Suppose G is not strictly increasing on $[\beta_0,\beta_1]$. Then, there exist $b_1, b_2 \in [\beta_0,\beta_1]$ such that $b_1 < b_2$ and $G(b_1) = G(b_2) = c$.

---

[11]In this and subsequent cases, differentiating under the integral sign is justified by our assumptions and Theorem XIII.8.1 in Lang [15].

As $b_1 < b_2 \le \beta_1 = \inf G^{-1}(\{1\})$, we have $c = G(b_1) < 1$. So, there exists $N \in \mathcal{N}$ such that $c + 1/N < 1$. For every $n \ge N$, we have $b(c + 1/n) \ge b_2 > b_1 \ge b(c)$. Thus, $b(c+) \ge b_2 > b(c)$, i.e., b is not continuous.

Conversely, suppose b is not continuous. As b is increasing and left-continuous, there exists $c \in [0,1)$ and $\alpha_1, \alpha_2 \in \Re$ such that $b(c) < \alpha_1 < \alpha_2 < b(c+) \le b(c + 1/n)$ for every $n \in \mathcal{N}$ such that $c + 1/n \le 1$. As G is increasing, we have $c \le G(\alpha_1) \le G(\alpha_2) < c + 1/n$. It follows that $G(\alpha_1) = c = G(\alpha_2)$. Thus, G is not strictly increasing.

Suppose b is not strictly increasing. Then, there exist $c_1, c_2 \in C$ such that $c_1 < c_2$ and $b(c_1) = b(c_2)$. Then, $G(b(c_1)) = G(b(c_2)) \ge c_2 > c_1$. If $c_1 = 0$, then $G(0) = G(b(0)) > 0 = G(0-)$ as $\beta_0 \ge 0$, i.e., G is not continuous. If $c_1 > 0$, then $G(b(c_1) - 1/n) < c_1$ for every $n \in \mathcal{N}$, which implies $G(b(c_1)) > c_1 \ge G(b(c_1)-)$, i.e., G is not continuous.

Conversely, suppose G is not continuous. As G is increasing and right-continuous, there exists $\alpha \in \Re$ and $c_1, c_2 \in (0,1)$ such that $G(\alpha) > c_2 > c_1 > G(\alpha-)$. As G is increasing, $[\alpha, \infty) \subset G^{-1}([c_1, 1])$. If $\beta < \alpha$, then $G(\beta) \le G(\alpha-) < c_1$, which means $\beta \notin G^{-1}([c_1, 1])$. So, $[\alpha, \infty) \supset G^{-1}([c_1, 1])$. Thus, $G^{-1}([c_1, 1]) = [\alpha, \infty)$. Similarly, $G^{-1}([c_2, 1]) = [\alpha, \infty)$. By definition, $b(c_1) = \alpha = b(c_2)$. Thus, b is not strictly increasing. ∎

Proof of Lemma 7.5 Suppose G is strictly increasing on $[\beta_0, \beta_1]$ and continuous. Then, Lemma 7.4 implies that b is continuous. Continuity of g follows immediately for the first-price auction. To confirm the continuity of g for the second-price auction, consider a sequence $(c_n) \subset C$ converging to c. Define $h : C \times \Re \to \Re$ by $h(c, y) = y 1_{[0, b(c)]}(y)$. As b is continuous, $\lim_{n\uparrow\infty} b(c_n) = b(c)$. Consequently, $\lim_{n\uparrow\infty} h(c_n, y) = \lim_{n\uparrow\infty} y 1_{[0, b(c_n)]}(y) = y 1_{[0, b(c)]}(y) = h(c, y)$ for $y \ne b(c)$. As G is continuous, $\mu_G(\{b(c)\}) = G(b(c)) - G(b(c)-) = 0$, where μ_G is the Lebesgue-Stieltjes measure generated by G. By definition, $g(.) = \int_{[0, b(.)]} y \, dG(y) = \int_{\Re} h(., y) \, dG(y) = \int_{\Re} \mu_G(dy) \, h(., y)$. Using the dominated convergence theorem (Billingsley [4], Theorem 16.4), it follows that $\lim_{n\uparrow\infty} g(c_n) = \lim_{n\uparrow\infty} \int_{\Re} \mu_G(dy) \, h(c_n, y) = \int_{\Re} \mu_G(dy) \, h(c, y) = g(c)$. ∎

Proof of Lemma 7.6 (A) Let $B(C)$ be the set of real-valued functions on C that are uniform limits of functions that are finite linear combinations of the characteristic functions of sets in the Borel σ-algebra $\mathcal{B}(C)$. With the supremum norm, $B(C)$ is a Banach space (Dunford and Schwartz [10], Section IV.5). As g is bounded and Borel measurable, $g \in B(C)$.

Let $\text{ba}(C)$ be the set of bounded, finitely additive real-valued functions defined on $\mathcal{B}(C)$. With the variation norm, $\text{ba}(C)$ is a Banach space (Dunford and Schwartz [10], Sections III.7 and IV.9). By a Riesz type representation theorem (Dunford and Schwartz [10], Theorem IV.5.1), $\text{ba}(C)$ may be identified with the conjugate of $B(C)$.

Equip $\text{ba}(C)$ with the $B(C)$ topology. This is the weakest topology on $\text{ba}(C)$ that makes the linear functional $\mu \mapsto \int_C \mu(dx) \, h(x)$ continuous for

every $h \in B(C)$. As $g \in B(C)$, L is continuous.

By Alaoglu's theorem (Dunford and Schwartz [10], Theorem V.4.2), the closed unit sphere of $\mathrm{ba}(C)$ is compact. As $\Delta(C, c)$ is a subset of this sphere, it is compact if it is closed. Consider $\lambda \in \mathrm{ba}(C)$ that is an accumulation point of $\Delta(C, c)$. Then, there exists a net $(\lambda_n) \subset \Delta(C, c)$ converging to λ. By definition, $\lim_n \lambda_n(E) = \lim_n \int_C \lambda_n(dx) \, 1_E(x) = \int_C \lambda(dx) \, 1_E(x) = \lambda(E)$ for every $E \in \mathcal{B}(C)$. Consequently, $\lambda(\emptyset) = 0$, $\lambda(C) = 1$ and $\lambda \geq 0$. Consider pairwise disjoint sets $E_1, \ldots, E_k \in \mathcal{B}(C)$. Then, $E = \cup_{i=1}^k E_i \in \mathcal{B}(C)$. As each λ_n is finitely additive, we have $\lambda(E) = \lim_n \lambda_n(E) = \lim_n \sum_{i=1}^k \lambda_n(E_i) = \sum_{i=1}^k \lim_n \lambda_n(E_i) = \sum_{i=1}^k \lambda(E_i)$. So, λ is finitely additive. As the identity map on C is bounded and measurable, it belongs to $B(C)$ and $c = \lim_n \int_C \lambda_n(dx) \, x = \int_C \lambda(dx) \, x$. Thus, $\lambda \in \Delta(C, c)$. So, $\Delta(C, c)$ is closed and therefore compact.

The compactness of $\Delta(C, c)$ and the continuity of L yield the result.

(B) As L is nonnegative and bounded, so is e.

If $c_1, c_2 \in C$ and $c_1 < c_2$, then $\{\lambda \in \Delta(C)^* \mid m_\lambda \geq c_2\} \subset \{\lambda \in \Delta(C)^* \mid m_\lambda \geq c_1\}$, and using Footnote 9, $e(c_2) = \inf\{L(\lambda) \mid \lambda \in \Delta(C)^* \land m_\lambda \geq c_2\} \geq \inf\{L(\lambda) \mid \lambda \in \Delta(C)^* \land m_\lambda \geq c_1\} = e(c_1)$.

Consider $c_1, c_2 \in C$, $t \in (0, 1)$ and $c = tc_1 + (1-t)c_2$. By (A), there exists $\lambda_1 \in \Delta(C, c_1)^*$ and $\lambda_2 \in \Delta(C, c_2)^*$ such that $e(c_1) = L(\lambda_1)$ and $e(c_2) = L(\lambda_2)$. Set $\lambda = t\lambda_1 + (1-t)\lambda_2$. Since $m_\lambda = tm_{\lambda_1} + (1-t)m_{\lambda_2} = tc_1 + (1-t)c_2 = c$, we have $\lambda \in \Delta(C, c)^*$. It follows that $e(c) \leq L(\lambda) = tL(\lambda_1) + (1-t)L(\lambda_2) = te(c_1) + (1-t)e(c_2)$. Thus, e is a convex function.

Continuity of e on $(0, 1)$ follows from Theorem 7 on page 193 of Berge [3].

The Busemann-Feller-Alexandrov theorem (see Fleming and Soner [11]) implies that e is twice differentiable Lebesgue almost everywhere. ∎

Proof of Lemma 7.7 The proof is by construction. Let $a = b - 2\epsilon$.

Define $\xi : \Re \to \Re$ by $\xi(x) = (2x - b - a)/(b - a)$. Clearly, ξ is $\mathcal{C}^\infty$, $\xi(a) = -1$, $\xi(b) = 1$, $|\xi(x)| < 1$ for $x \in (a, b)$ and $|\xi(x)| > 1$ for $x \in \Re \setminus [a, b]$.

Define $\psi : \Re \to \Re$ by[12]

$$\psi(x) = \begin{cases} \exp\{-(1 - x^2)^{-1}\}, & \text{if } |x| < 1 \\ 0, & \text{if } |x| \geq 1 \end{cases}$$

Clearly, $\mathrm{supp}\, \psi = [-1, 1]$ and $\int_\Re dx \, \psi \circ \xi(x) = \int_{[a,b]} dx \, \psi \circ \xi(x) \leq \epsilon$. We shall verify that ψ is $\mathcal{C}^\infty$.

Define $f : \Re \to \Re$ by $f(x) = -\psi \circ \xi(x) / \int_\Re dx \, \psi \circ \xi(x)$. It is easily verified that $f \leq 0$, $\mathrm{supp}\, f = [a, b]$, f is symmetric around $(a + b)/2$ and $\int_\Re dx \, f(x) = \int_{[a,b]} dx \, f(x) = -1$. As $-f$ is a probability density function on $\Re$, with support $[a, b]$ and symmetric around $(a + b)/2$, it follows that $\int_\Re dx \, x f(x) = \int_{[a,b]} dx \, x f(x) = -(a + b)/2$. As ξ and ψ are $\mathcal{C}^\infty$, so is f.

---

[12]Such "bump functions" are widely used in analysis, for instance as mollifiers and to construct smooth partitions of unity.

Define $g : \Re \to \Re$ by $g(x) = 1 + \int_{(-\infty, x]} dy\, f(y)$. Then, g is $\mathcal{C}^\infty$, $Dg = f \leq 0$, $g = 1$ on $(-\infty, a]$, $g \in (0,1)$ on (a,b) and $g = 0$ on $[b, \infty)$.

Define $h : \Re \to \Re$ by

$$h(x) = \begin{cases} x - b + \epsilon, & \text{if } x < a \\ a - b + \epsilon + \int_{[a,x]} dy\, g(y), & \text{if } x \geq a \end{cases}$$

It is easily verified that h is $\mathcal{C}^\infty$, $Dh = g \geq 0$ and $D^2 h = f \leq 0$. Thus, h is increasing and concave, with $Dh = 1$ on $(-\infty, a]$, $Dh \in (0,1)$ on (a,b) and $Dh = 0$ on $[b, \infty)$. We now compare h and u.

Consider $x \in (-\infty, a)$. We have $h(x) = x - b + \epsilon = u(x) + \epsilon$.

Consider $x \in (b, \infty)$. As $g = 0$ on (b, ∞), $\int_{[a,x]} dy\, g(y) = \int_{[a,b]} dy\, g(y) = \int_{[a,b]} dy\, [1 + \int_{[a,y]} dz\, f(z)] = b - a + b \int_{[a,b]} dz\, f(z) - \int_{[a,b]} dy\, y f(y) = (b-a)/2 = \epsilon$. So, $h(x) = a - b + 2\epsilon = 0 = u(x)$.

Consider $x \in [a,b]$. As $f = 0$ on $(-\infty, a]$, $\int_{[a,x]} dy\, g(y) = \int_{[a,x]} dy\, [1 + \int_{[a,y]} dz\, f(z)] = x - a + \int_{[a,x]} dy \int_{[a,y]} dz\, f(z)$. As $f \leq 0$, we have $0 \geq \int_{[a,x]} dy \int_{[a,y]} dz\, f(z) = \int_{[a,x]} dy\, (x - y) f(y) \geq \int_{[a,b]} dy\, (b - y) f(y) = (a - b)/2 = -\epsilon$. Therefore, $h(x) = a - b + \epsilon + x - a + \int_{[a,x]} dy \int_{[a,y]} dz\, f(z) = u(x) + \epsilon + \int_{[a,x]} dy \int_{[a,y]} dz\, f(z)$. Since $\int_{[a,x]} dy \int_{[a,y]} dz\, f(z) \in [-\epsilon, 0]$, we have $h(x) - u(x) \in [0, \epsilon]$.

Consequently, $h(b) = 0$, $h \geq u$ and $\sup\{|h(x) - u(x)| \mid x \in \Re\} \leq \epsilon$.

It only remains to confirm that ψ is $\mathcal{C}^\infty$. Observe that $\psi = \phi \circ \eta$, where $\phi : \Re \to \Re$ is given by

$$\phi(y) = \begin{cases} e^{-1/y}, & \text{if } y > 0 \\ 0, & \text{if } y \leq 0 \end{cases} \tag{4}$$

and η is given by $\eta(x) = 1 - x^2$. Using the chain-rule, as η is $\mathcal{C}^\infty$, it suffices to show that ϕ is $\mathcal{C}^\infty$. As preparation, we note two facts.

First, for $n \in \mathcal{N}$, it is easily verified that

$$D^n \phi(y) = \begin{cases} p_n(y)\phi(y) y^{-2n}, & \text{if } y > 0 \\ 0, & \text{if } y < 0 \end{cases} \tag{5}$$

where $p_n(y)$ is the polynomial of degree $n - 1$ generated by the recursive rule: $p_1(y) = 1$ and $p_{k+1}(y) = y^2 Dp_k(y) - (2ky - 1)p_k(y)$ for $k \in \mathcal{N}$.

Second, consider $y > 0$ and $n \in \mathcal{N} \cup \{0\}$. Using the power series representation of $e^{1/y}$, we have the estimate $0 \leq y^{-n} = yy^{-(n+1)} \leq y(n+1)! e^{1/y}$. Thus, $0 \leq \lim_{y \downarrow 0} e^{-1/y} y^{-n} \leq \lim_{y \downarrow 0} y(n+1)! = 0$, i.e., $\lim_{y \downarrow 0} \phi(y) y^{-n} = 0$.

Setting $n = 0$ implies that $\lim_{y \downarrow 0} \phi(y) = 0$. Using (4), ϕ is continuous.

The left-hand derivative of ϕ at 0 is $\lim_{y \uparrow 0} [\phi(y) - \phi(0)]/y = 0$ and the right-hand derivative of ϕ at 0 is $\lim_{y \downarrow 0} [\phi(y) - \phi(0)]/y = \lim_{y \downarrow 0} \phi(y) y^{-1} = 0$. Thus, $D^1 \phi(0) = 0$. Using (5), $D^1 \phi$ exists on $\Re$ and is continuous.

For the inductive step, suppose $D^n \phi$ exists on $\Re$ for $n \in \mathcal{N}$ and $D^n \phi(0) = 0$. The left-hand derivative of $D^n \phi$ at 0 is $\lim_{y \uparrow 0} [D^n \phi(y) - D^n \phi(0)]/y = 0$

and the right-hand derivative of $D^n\phi$ at 0 is $\lim_{y\downarrow 0}[D^n\phi(y) - D^n\phi(0)]/y = \lim_{y\downarrow 0} p_n(y)\phi(y)y^{-(2n+1)} = p_n(0)\lim_{y\downarrow 0}\phi(y)y^{-(2n+1)} = 0$. It follows that $D^{n+1}\phi(0) = 0$. Using (5), $D^{n+1}\phi$ exists on $\Re$ and is continuous.

It follows that ϕ is $\mathcal{C}^\infty$. ■

Proof of Lemma 7.8 Let $Y = [-\gamma, 0] \times [a, b]$ and define $A : Y \to \Re$ by $A(v, c) = -v - c$. Endow $\mathcal{C}(Y)$ and $\mathcal{C}(A(Y))$ with their respective compact-open topologies. As Y and $A(Y)$ are compact, these topologies are metrized by the respective uniform metrics (Dugundji [9], Theorem XII.8.2).

Since the sequence $(g_n) \subset \mathcal{C}(\Re)$ converges uniformly to g, the sequence $(I - g_n) \subset \mathcal{C}(\Re)$ converges uniformly to $I - g$. Therefore, $(I - g_n)$ converges uniformly to $I - g$ on $A(Y)$. As the mapping $f \mapsto f \circ A$ from $\mathcal{C}(A(Y))$ to $\mathcal{C}(Y)$ is continuous (Dugundji [9], Theorem XII.2.1), the sequence $((I - g_n) \circ A) \subset \mathcal{C}(Y)$ converges uniformly to $(I - g) \circ A$ on Y.

As the mapping $h \mapsto P \circ h$ from $\mathcal{C}(Y)$ to $\mathcal{C}(Y)$ is continuous (Dugundji [9], Theorem XII.2.1), the sequence $(P \circ (I - g_n) \circ A) \subset \mathcal{C}(Y)$ converges uniformly to $P \circ (I - g) \circ A$ on Y.

Since $u_n = -S \circ \pi_2 - P \circ (I - g_n) \circ A$, where π_2 is the projection $\pi_2(v, c) = c$, it follows that the sequence (u_n) converges uniformly to $u = -S \circ \pi_2 - P \circ (I - g) \circ A$ on Y. The claim follows as $v(O) \times C \subset Y$. ■

Proof of Lemma 7.9 Fix $n \geq N$. It is easily confirmed that $D_{22}u_n < 0$. As $2/n < a$ and $Dg_n(x) = 1$ for $x \in (-\infty, -2/n]$, we have $Dg_n(x) = 1$ for every $x \in [-b, -a]$. Since $0 \in O$, $v_1 = \sup v(O) = 0$. Consequently, for every $c \in C$, $D_1u_n(v_1, c) = D_1u_n(0, c) = DP((I - g_n)(-c))[1 - Dg_n(-c)] = 0$. Routine calculation yields $D_{211}u_n = -DPD^3g_n - 3D^2P(1 - Dg_n)D^2g_n + (1 - Dg_n)^3D^3P$, where the derivatives of P are evaluated at $(I - g_n)(w - c)$ and the derivatives of g_n are evaluated at $w - c$. If $w - c < 0$, then $(I - g_n)(w - c) = w - c - g_n(w - c) < w - c - \min\{w - c, 0\} = 0$. Consequently, $DP \circ (I - g_n)(w - c) = 0$ as $P(x) = 0$ for $x \leq 0$. On the other hand, if $w - c \geq 0$, then $D^3g_n(w - c) = 0$ as $g_n(x) = 0$ for $x \geq 0$. Given our hypotheses regarding P and the facts that g_n is concave and $Dg_n \leq 1$, we have $D_{211}u_n = -3D^2P(1 - Dg_n)D^2g_n + (1 - Dg_n)^3D^3P \geq 0$. By Theorem 7.1, $Dc(r; n) \geq 0$. ■

Proof of Lemma 7.10 Fix $r \in \{r_1, r_2\}$ and the subsequence $(c(r, n_j))$ of $(c(r, n))$, converging to $c^*(r)$. For $m, j \in \mathcal{N}$, let $F(m, j) = U(r, c(r, n_m); n_j)$, $f(m) = U(r, c(r, n_m))$ and $\phi(j) = U(r, c^*(r); n_j)$. For $(m_1, j_1), (m_2, j_2) \in \mathcal{N}^2$, we say that $(m_1, j_1) \succeq (m_2, j_2)$ if $m_1 \geq m_2$ and $j_1 \geq j_2$. Then, $(\mathcal{N}^2, \succeq)$ is a directed set and $\{F(m, j) \mid (m, j) \in \mathcal{N}^2; \succeq\}$ is a net in $\Re$.

Fix $\epsilon > 0$. By Lemma 7.8, (u_n) converges uniformly to u on $v(O) \times C$. Therefore, (u_{n_j}) converges uniformly to u on $v(O) \times C$. So, there exists $J \in \mathcal{N}$ such that $j \geq J$ implies $\|u_{n_j} - u\| < \epsilon/8$, where $\|.\|$ is the supremum norm on $\mathcal{C}(v(O) \times C)$. Consequently, for every $c \in C$, $j \geq J$ implies

$$|U(r, c; n_j) - U(r, c)| \leq \int_O \mu(r)(dx) |u_{n_j}(v(x), c) - u(v(x), c)|$$

$$\leq \|u_{n_j} - u\|$$
$$< \epsilon/8 \tag{6}$$

Thus, $\lim_{j\uparrow\infty} U(r,c;n_j) = U(r,c)$ for every $c \in C$.

Setting $c = c^*(r)$ in (6), we have $|\phi(j) - U(r,c^*(r))| = |U(r,c^*(r);n_j) - U(r,c^*(r))| < \epsilon/8$ for $j \geq J$. Thus, $\lim_{j\uparrow\infty} \phi(j) = U(r,c^*(r))$.

If we set $c = c(r,n_m)$ in (6), then $m \in \mathcal{N}$ and $j \geq J$ imply $|F(m,j) - f(m)| = |U(r,c(r,n_m);n_j) - U(r,c(r,n_m))| < \epsilon/8$. Consequently, $m \in \mathcal{N}$ and $j \geq J$ implies $|F(m,j) - F(m,J)| \leq |F(m,j) - f(m)| + |F(m,J) - f(m)| < \epsilon/4$.

The continuity of u_{n_j} and the bounded convergence theorem (Billingsley [4], Theorem 16.5) imply that $\phi(j) = \lim_{m\uparrow\infty} F(m,j)$ for every $j \in \mathcal{N}$. So, given J, there exists $M \in \mathcal{N}$ such that $m \geq M$ implies $|F(m,J) - \phi(J)| < \epsilon/8$. Consequently, $m \geq M$ implies $|F(m,J) - F(M,J)| \leq |F(m,J) - \phi(J)| + |F(M,J) - \phi(J)| < \epsilon/4$.

It follows that $j \geq J$ and $m \geq M$ implies $|F(m,j) - F(M,J)| \leq |F(m,j) - F(m,J)| + |F(m,J) - F(M,J)| < \epsilon/2$. Therefore, if $(m_1,j_1) \succeq (M,J)$ and $(m_2,j_2) \succeq (M,J)$, then $|F(m_1,j_1) - F(m_2,j_2)| \leq |F(m_1,j_1) - F(M,J)| + |F(m_2,j_2) - F(M,J)| < \epsilon$. Thus, $\{F(m,j) \mid (m,j) \in \mathcal{N}^2; \succeq\}$ is a Cauchy net with respect to the Euclidean space $\Re$. As the Euclidean space $\Re$ is complete, this net converges to some $p \in \Re$ (Dugundji [9], Theorem XIV.3.2). Consequently, there exists $(M',J') \in \mathcal{N}^2$ such that $(m,j) \succeq (M',J')$ implies $|F(m,j) - p| < \epsilon$.

Without loss of generality, let $J' \geq M'$. So, $j \geq J'$ implies $(j,j) \succeq (J',J') \succeq (M',J')$, which implies $|F(j,j) - p| < \epsilon$. So, $\lim_{j\uparrow\infty} F(j,j) = p$.

Moreover, $|p - \phi(j)| = |p - \lim_{m\uparrow\infty} F(m,j)| = \lim_{m\uparrow\infty} |p - F(m,j)| \leq \epsilon$ for every $j \geq J'$. Thus, $\lim_{j\uparrow\infty} \phi(j) = p$.

By definition, $U(r,c(r,n_j);n_j) \geq U(r,c;n_j)$ for all $j \in \mathcal{N}$ and $c \in C$. Therefore, we have $U(r,c^*(r)) = \lim_{j\uparrow\infty} \phi(j) = p = \lim_{j\uparrow\infty} F(j,j) = \lim_{j\uparrow\infty} U(r,c(r,n_j);n_j) \geq \lim_{j\uparrow\infty} U(r,c;n_j) = U(r,c)$ for every $c \in C$. ∎

Proof of Lemma 7.14 If v is continuous, then it is Borel measurable, and compactness of O implies it is bounded. So, it suffices to show that v is continuous and concave in each case.

In Example 7.11, for every $i \in I$, as the projection mapping $\pi_i : X \to \Re^l$ is continuous, $f_i \circ \pi_i : O \to \Re$ is continuous. Since $v(.) = \min\{f_i \circ \pi_i(.) \mid i \in I\}$, v is continuous (Berge [3], Theorems IV.8.3 and IV.8.4).

Now consider $x,y \in O$ and $t \in (0,1)$. Then, $tx + (1-t)y \in O$ and v is concave as

$$
\begin{aligned}
v(tx + (1-t)y) &= \min\{f_i(tx_i + (1-t)y_i) \mid i \in I\} \\
&\geq \min\{tf_i(x_i) + (1-t)f(y_i) \mid i \in I\} \\
&\geq \min\{tf_i(x_i) \mid i \in I\} + \min\{(1-t)f_i(y_i) \mid i \in I\} \\
&= t\min\{f_i(x_i) \mid i \in I\} + (1-t)\min\{f_i(y_i) \mid i \in I\}
\end{aligned}
$$

$$= tv(x) + (1-t)v(y) \tag{7}$$

Now consider Example 7.12. Clearly, S has nonempty values. Consider a sequence $(x_n)_{n\in\mathcal{N}} \subset O$ and a family of sequences $\{(x_n^i)_{n\in\mathcal{N}} \subset \Re_+^l \mid i \in I\}$ such that $\lim_n x_n = x$, $\lim_n x_n^i = x^i$ for every $i \in I$ and $x_n = \sum_{i\in I} x_n^i$ for every $n \in \mathcal{N}$. Clearly, $(x_n^i)_{i\in I} \in S(x_n)$ for every $n \in \mathcal{N}$. It follows that $x \in O$, $x^i \in \Re_+^l$ for every $i \in I$ and $x = \sum_{i\in I} x^i$. It follows that $(x^i)_{i\in I} \in S(x)$. Thus, S has a closed graph, and therefore, closed values. As O is compact, there exists $y \in (\Re_+^l)^I$ such that $S(x)$ is a subset of the compact set $\{z \in (\Re^l)^I \mid 0 \le z \le y\}$ for every $x \in O$. Therefore, S is upper hemicontinuous (Berge [3], Theorem VI.1.7) and has compact values.

We now confirm that S is lower hemicontinuous. Fix $x \in O$ and $(x^i)_{i\in I} \in S(x)$. By definition, $0 \ll x = \sum_{i\in I} x^i$; therefore, $\pi_j(x) > 0$ for every $j \in \{1,\ldots,l\}$. Consider a sequence $(x_n)_{n\in\mathcal{N}} \subset O$ converging to x. We need to construct a family of sequences $\{(x_n^i)_{n\in\mathcal{N}} \subset \Re_+^l \mid i \in I\}$ such that (a) $\sum_{i\in I} x_n^i = x_n$ for every $n \in \mathcal{N}$, and (b) $\lim_n x_n^i = x^i$ for every $i \in I$.

For $j \in \{1,\ldots,l\}$ and $n \in \mathcal{N}$, let $\lambda_n^j = \pi_j(x_n)/\pi_j(x)$. As $\lim_n x_n = x$, we have $\lim_n \pi_j(x_n) = \pi_j(x)$ and $\lim_n \lambda_n^j = 1$ for every $j \in \{1,\ldots,l\}$.

Let $x_n^i = (\lambda_n^1 \pi_1(x^i),\ldots,\lambda_n^l \pi_l(x^i))$. Then, $x_n^i \in \Re_+^l$ and $\lim_n x_n^i = x^i$ for every $i \in I$. For $j \in \{1,\ldots,l\}$, we have $\sum_{i\in I} \lambda_n^j \pi_j(x^i) = \lambda_n^j \pi_j(\sum_{i\in I} x^i) = \lambda_n^j \pi_j(x)$. It follows that $\sum_{i\in I} x_n^i = (\sum_{i\in I} \lambda_n^1 \pi_1(x^i),\ldots,\sum_{i\in I} \lambda_n^l \pi_l(x^i)) = (\lambda_n^1 \pi_1(x),\ldots,\lambda_n^l \pi_l(x)) = (\pi_1(x_n),\ldots,\pi_l(x_n)) = x_n$, as required.

As S is continuous and has nonempty compact values, v is continuous (Berge [3], Maximum theorem, Pg. 116).

Consider $x,y \in O$ and $t \in (0,1)$. Then, $tx + (1-t)y \in O$. Using (A) and the fact that S has nonempty compact values, we conclude that $v(x) = \min\{f_i(x_i) \mid i \in I\}$, $v(y) = \min\{f_i(y_i) \mid i \in I\}$ and $v(tx + (1-t)y) = \min\{f_i(z_i) \mid i \in I\}$ for some $(x_i)_{i\in I} \in S(x)$, $(y_i)_{i\in I} \in S(y)$ and $(z_i)_{i\in I} \in S(tx + (1-t)y)$. Since $(tx_i + (1-t)y_i)_{i\in I} \in S(tx + (1-t)y)$, (7) implies that v is concave as

$$
\begin{aligned}
v(tx + (1-t)y) &= \min\{f_i(z_i) \mid i \in I\} \\
&\ge \min\{f_i(tx_i + (1-t)y_i) \mid i \in I\} \\
&\ge tv(x) + (1-t)v(y)
\end{aligned}
$$

In Example 7.13, v is continuous and concave by an argument analogous to that used for Example 7.11. ∎

References

[1] K. J. Arrow: *Essays in the Theory of Risk-bearing*, North-Holland Publishing Company (1970)

[2] R. Bartle: *The Elements of Real Analysis*, Wiley (1976)

[3] C. Berge: *Topological Spaces*, Macmillan (1963)

[4] P. Billingsley: *Probability and Measure, 3rd ed*, Wiley Interscience (1995)

[5] D. Blackwell: Comparison of experiments, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 93-102 (1951)

[6] D. Blackwell: Equivalent comparisons of experiments, *The Annals of Mathematical Statistics* 24, 265-272 (1953)

[7] P. Cartier, J. M. G. Fell, P. A. Meyer: Comparaison des mesures portées par un ensemble convexe compact, *Bulletin de la Société Mathématique de France* 92, 435-445 (1964)

[8] P. Diamond, J. Stiglitz: Increases in risk and in risk aversion, *Journal of Economic Theory* 8, 337-360 (1974)

[9] J. Dugundji: *Topology*, Wm. C. Brown Publishers (1989)

[10] N. Dunford, J. Schwartz: *Linear Operators, Part I: General Theory*, Wiley Interscience (1988)

[11] W. Fleming, H. Soner: *Controlled Markov Processes and Viscosity Solutions*, Springer-Verlag (1993)

[12] G. H. Hardy, J. E. Littlewood, G. Pólya: *Inequalities, 2nd ed.*, Cambridge University Press (1988)

[13] J. M. Howie: *Fundamentals of Semigroup Theory*, Clarendon Press (1996)

[14] J.-J. Laffont, J. Tirole: Using cost observation to regulate firms, *Journal of Political Economy* 94, 614-641 (1986)

[15] S. Lang: *Real and Functional Analysis*, Springer-Verlag (1993)

[16] L. Le Cam: Comparison of experiments - A short review, in T. S. Ferguson, L. S. Shapley, J. B. MacQueen (eds.): *Statistics, Probability and Game Theory: Papers in Honor of David Blackwell*, IMS Lecture Notes-Monograph Series 30, 127-138 (1996)

[17] K. Mosler, M. Scarsini: Some theory of stochastic dominance. In K. Mosler, M. Scarsini (ed.) *Stochastic Orders and Decision under Risk*, IMS Lecture Notes-Monograph Series 19 (1991)

[18] K. Parthasarathy: *Probability Measures on Metric Spaces*, Academic Press (1967)

[19] M. Perlman: Jensen's inequality for a convex vector-valued function on an infinite-dimensional space, *Journal of Multivariate Analysis* 4, 52-65 (1974)

[20] B. Pettis: On integration in vector spaces, *Transactions of the American Mathematical Society* 44, 277-304 (1938)

[21] J. W. Pratt: Risk aversion in the small and in the large, *Econometrica* 32, 122-136 (1964)

[22] M. Rothschild, J. Stiglitz: Increasing risk: I. A definition, *Journal of Economic Theory* 2, 225-243 (1970)

[23] W. Russell, T. Seo: Ordering uncertain prospects: the multivariate utility functions case, *The Review of Economic Studies* 45, 605-610 (1978)

[24] S. Shah: Comparative aversion to vector-valued risks, available at `http://ssrn.com/abstract=1155470` (2013)

[25] M. Shaked, J. Shanthikumar: *Stochastic Orders*, Springer (2007)

[26] S. Sherman: On a theorem of Hardy, Littlewood, Pólya and Blackwell, *Proceedings of the National Academy of Sciences* 37, 826-831 (1951)

[27] C. Stein: *Notes on a Seminar on Theoretical Statistics. I. Comparison of experiments*, Report, University of Chicago (1951)

[28] V. Strassen: The existence of probability measures with given marginals, *The Annals of Mathematical Statistics* 36, 423-439 (1965)

[29] E. N. Torgersen: *Comparison of Statistical Experiments*, Cambridge University Press (1991)