MGT-621 Microeconomics

Risk and Uncertainty

Thomas A. Weber*

February 2016

1 Introduction

There is a longstanding tradition, commonly attributed to Frank Knight (1921, pp. 19–20, 199–232), to distinguish between risk and uncertainty. After a whole chapter on the two notions, their differences and commonalities, Knight concludes (on page 233):

> [t]he practical difference between the two categories, risk and uncertainty, is that in the former the distribution of the outcome in a group of instances is known (...), while in the case of uncertainty this is not true, the reason being in general that it is impossible to form a group of instances, because the situation dealt with is in a high degree unique.

Just preceding that paragraph he noted,

> [w]e can also employ the terms "objective" and "subjective" probability to designate the risk and uncertainty respectively, as these expressions are already in general use with a signification akin to that proposed.

In Knight's distinction *risk* corresponds to an *objective* notion of event probability which is (at least in principle) verifiable, for instance by experiment.

*Chair of Operations, Economics and Strategy, École Polytechnique Fédérale de Lausanne, Station 5, CH-1015 Lausanne, Switzerland. Phone: +41 (21) 693 01 41. E-mail: thomas.weber@epfl.ch.

Thus, you might agree that the probability of an unweighted coin to show "tail" after a single toss is approximately 1/2. On the other hand, if you think about the event of Stanford winning the next "Big Game" in football against Berkeley (Cal) the *subjective* assessment of the probability of a Stanford win might be quite different for each one of you. The event would thus – in the Knightian distinction – fall under *uncertainty*. We will see, however, that given a rather small set of assumptions on your preferences over lotteries, it is possible to *construct* a unique "subjective" probability distribution over an exhaustive, mutually exclusive list of events ({Stanford wins, Stanford loses}) dissolving the material difference between risk and uncertainty. In the remainder of this course we will therefore simply disregard the distinction between risk and uncertainty and use both terms interchangeably. We will also assume that a probability distribution of the uncertain decision-relevant events is given, be it either objectively or constructed by evaluating subjective preferences over lotteries of outcomes.

If there is no fundamental difference between risk and uncertainty, one might still be able to argue about the *strength of beliefs*: for instance, we can be *pretty* (though maybe not absolutely) certain that the probability of "tail" in a single-coin toss is 1/2, whereas you might not have a lot of confidence in your subjective assessment that with 75 percent probability Stanford will win the next Big Game. Note though, that a lack of confidence cannot have any effect on any *terminal* action that you need to take based on a subjective probability assessment. Thus, as an example, if you were forced to make a fair bet of 10 dollars you should be indifferent between betting the money on Stanford or on the event ("tail","tail") after tossing two coins simultaneously.[1] On the other hand, if the terminal action (making the bet) is not immediately required, a lack of confidence might provoke you to take an *informational* action (e.g., you could find out more about the two teams and/or consult experts) *before* you commit to a terminal action.

Overview. In this lecture we will formulate the expected utility paradigm, show how representations of utility functions are uniquely (up to positive

[1]There might be behavioral reasons why you might still prefer the coin toss, such as "ambiguity aversion": individuals often prefer situations where event probabilities are well known to situations in which the likelihood of the different outcomes is ambiguous. An axiomatic explanation of ambiguity aversion (based on a relaxation of the independence axiom by Von Neumann and Morgenstern, cf. Section 3) has been proposed by Schmeidler (1989).

linear transformations) implied by a set of simple choice axioms, and how risks, after they have been quantified, can be ordered with respect to different classes of utility functions. We conclude this lecture with some remarks on comparative statics under uncertainty and a few applications.

2 The Expected Utility Paradigm

Every decision making problem consists in selecting a preferred action out of a (closed) set of possible actions $\mathcal{A}$, given a set of possible consequences $\mathcal{C}$. The occurrence of a particular consequence $c(a, \omega)$ depends – in addition to the decision maker's *action* a – on an uncertain *state of nature* ω which is an element of the *state space* Ω. Consequences are related to actions and events by a *consequence function* $c : \mathcal{A} \times \Omega \to \mathcal{C}$. The decision maker evaluates a particular consequence using a scalar (measurable) *utility function* $u : \mathcal{C} \to \mathbb{R}$ that assigns a value to its desirability.[2] The *expected utility* of an action $a \in \mathcal{A}$ is

$$EU(a) = \int_{\Omega} u(c(a, \omega)) d\mu(\omega), \tag{1}$$

where μ is an appropriate probability measure defined on (the Lebesgue-completion of) $\mathcal{R}$. Representation (1) is very general; depending on the circumstances it may however be useful to vary it somewhat. We will now briefly discuss two variations.

1. *Countable Partition of* Ω. A collection of *events* $\{\mathcal{E}_k\}_{k=1}^{\infty}$ which are pairwise disjoint (i.e., $E_k \cap E_j = \emptyset$ for any $k \neq j$) and exhaustive (i.e., $\bigcup_{k=1}^{\infty} \mathcal{E}_k = \Omega$), is called a *partition* of Ω.[3] If on any such partition

[2] Hirshleifer and Riley (1992, pp. 13–21) distinguish between a *preference-scaling function* (also called "elementary utility function") which measures the desirability of consequences and a *utility function* which measures the desirability of actions. We will call *utility function* a measure of the desirability of consequences (which depend on actions and events) and *expected utility* the measure of the desirability of an action.

[3] Strictly speaking, each *event* $\mathcal{E}_k$ is an element of a σ-algebra $\mathcal{R} \subset P(\Omega)$, whereby $P(\Omega)$ denotes the collection of all subsets of Ω. The decision maker assesses the probability of $\mathcal{E}_k$ using a probability measure $\mu(\mathcal{E}_k)$. The σ-algebra $\mathcal{R}$ is a family of subsets of Ω that contains Ω and is closed under the operations of countable union, intersection, and difference. A probability measure μ is a real nonnegative function defined on $\mathcal{R}$ such that μ is countably additive (for any countable number of disjoint subsets $\mathcal{E}_1, \mathcal{E}_2, \ldots$ of Ω: $\mu(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \ldots) = \mu(\mathcal{E}_1) + \mu(\mathcal{E}_2) + \cdots$) and $\mu(\Omega) = 1$ (normalization). The triplet $(\Omega, \mathcal{R}, \mu)$ is referred to as a *probability space*. For a short and rigorous exposition of measure theory see Kirillov and Gvishiani (1982, pp. 12–37). A classical treatment is by Halmos (1950).

consequences only depend on an action and on the particular event (and not on the particular state of nature contained in an event), i.e., if $c(a, \omega_k) = c_k(a)$ for all $\omega_k \in \mathcal{E}_k$, $k = 1, 2, \ldots$, then representation (1) is equivalent to

$$EU(a) = \sum_{k=1}^{\infty} \mu(\mathcal{E}_k) u(c_k(a)). \tag{2}$$

We can therefore replace the state space Ω with the partition of Ω, reducing the cardinality of the state space and possibly its dimensionality. In other words, each event $\mathcal{E}_k$ in the partition can be seen as an elementary event by itself and it is therefore unnecessary to distinguish between different elements $\omega \in \mathcal{E}_k$. By construction, changing from (1) to (2) does not affect the decision maker's preferences over actions.

2. *Decomposable Consequences,* $\mathcal{C} = \mathcal{A} \times \mathcal{X} \times \Omega$. If the set of consequences can be represented as the cartesian product of the action space $\mathcal{A}$, a cartesian outcome space $\mathcal{X} \subset \mathbb{R}^m$ (for some $m \geq 1$) and the state space Ω, then it is possible to rewrite (1),

$$EU(a) = \int_{\Omega} \left(\int_{\mathcal{X}} u(a, x, \omega) d\nu(x; a, \omega) \right) d\mu(\omega), \tag{3}$$

where ν is a probability measure (actually the cdf) of the random variable $\tilde{x} \in \mathcal{X}$ of the form $\text{Prob}(\tilde{x} < x | a, \omega) = \nu(x; a, \omega)$. The set of consequences $\mathcal{C}$ is decomposable in this manner, if by selecting an action $a \in \mathcal{A}$ the decision maker chooses a probability distribution of possible outcomes in $\mathcal{X}$.

The following examples outline how an appropriate expression for the expected utility can be obtained in practical settings.

Example 1 Consider the *standard portfolio problem* in which an investor with initial wealth w selects how much she should invest in a risky asset of uncertain return $\tilde{r} \in [-1, 1] = \Omega$ and how much she should hold in cash at zero return. In this decision making problem the investor's action a is the amount that she invests in the risky asset: in the absence of borrowing, $a \in [0, w] = \mathcal{A}$. The consequence of any action a also depends on the uncertain return $\tilde{r}$, so that $c(a, \tilde{r}) = w + a\tilde{r}$. Assuming that the investor always strictly prefers more wealth, her utility function u evaluating the consequences is strictly increasing. It is convenient to assume that her utility function is also continuous.[4] Let us now construct a probability measure for the interesting events $\{r\}$, $r \in \Omega$. Note first that these events are disjoint but not countable, so that our expected utility formula (1) does *not* immediately apply. In fact the σ-algebra $\mathcal{R}$ needed to accommodate all interesting events is far larger than their union. The construction of a set-measure μ on $\mathcal{R}$ proceeds as follows. First, one can verify that the σ-algebra $\mathcal{R}$ generated by the relevant events $\{r\}$ is (in addition to containing Ω itself) spanned by the collection of all half-open intervals $[a, b) \subset \Omega$. Second, by additivity it is $\mu([a, b)) = \mu([-1, b)) - \mu([-1, a))$, so that with the "cumulative distribution function" (cdf) $F(x) = \mu([-1, x)) = \text{Prob}(\tilde{r} < x)$ any set in $\mathcal{R}$ can be measured in terms of F. Using the set-measure μ it is possible to construct a Lebesgue integral which allows computing the weighted sum (1), even if the underlying state space is not countable (but is merely Lebesgue-measurable):

$$EU(a; w) = \int_{-1}^{1} u(w + ar) dF(r).$$

The cdf F represents the decision maker's beliefs about the distribution of outcomes and needs to be specified in order to determine a solution to the decision making problem.[5] One can show that since u is continuous $EU(\cdot)$ is also continuous. Since the action set $\mathcal{A}$ is compact, an optimal action a^* maximizes her expected utility, and a solution to her decision problem consists thus in finding

$$a^*(w) \in \max_{a \in \mathcal{A}} EU(a; w).$$

We will revisit the standard portfolio problem later in this course. $\square$

Example 2 Consider a monopolistic firm that faces the decision ($a \in [0, \infty)$) to enter an uncertain market ($a > 0$) with a new product or not entering

[4]Proposition 1 provides a necessary and sufficient condition for a utility representation of preferences. If preferences are continuous (which means that they are preserved under limits), then the utility representation is also continuous (Mas-Colell et al., 1995, pp. 46–47).

[5]The beliefs F could be *either* exogenously provided as an *objective* representation of a cdf, *or* it could be endogenously constructed based on the decision maker's preferences over lotteries. If F is differentiable (almost everywhere), then a *probability density function* (pdf) $f = F'$ can be obtained and the resulting Riemann-Stieltjes integral is computed by evaluating the following ordinary Riemann integral: $EU(a) = \int_{-1}^{1} c(a, r) f(r) dr$.

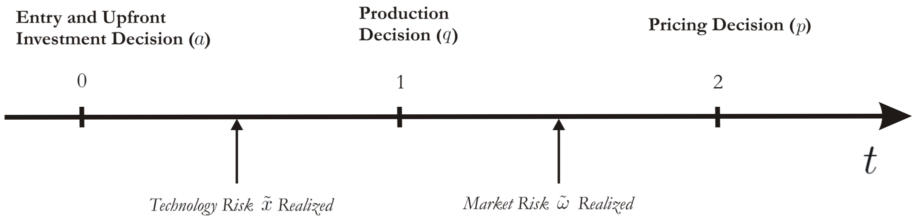

Figure 1: Timeline for the sequential decision problem in Example 2.

($a = 0$). Entering the market involves making an *irreversible* upfront investment $ka > 0$, where k is a positive constant representing the "capital intensity" of the production process. There is uncertainty about both the demand for the new product ("market risk") and the efficiency of the production process ("technology risk") which is still to be developed. For simplicity we assume that the two types of risks are *independent*. To represent the market risk assume that the firm faces a linear demand curve with unknown intercept, i.e., at price p it faces a random demand $\tilde{D}(p) = \tilde{\omega} - p$, where $\tilde{\omega}$ is distributed on $\Omega = \mathbb{R}_+$ with cdf F. The technology risk, which can be influenced by the firm through its investment decision, is represented by a random marginal cost $\tilde{x}/a$, whereby $\tilde{x}$ is distributed on $\mathcal{X} = \mathbb{R}_+$ with cdf G. Increasing a thus in expectation increases productive efficiency by lowering marginal cost. The firm has to make three decisions in three time periods $t \in \{0, 1, 2\}$. At time $t = 0$, it decides about its upfront investment ka. Subsequently the technological uncertainty $\tilde{x}$ realizes, and at time $t = 1$, the firm fixes its production quantity $q^*(a, x)$. At time $t = 2$, the demand realizes and the firm sets its price p^* (cf. the timeline in Figure 1). To solve such a dynamic decision making problem under uncertainty it is often useful to adopt a *dynamic programming* approach by starting the solution at the last period working backwards. At time $t = 2$, the firm's (deterministic) profits can be written as

$$\Pi_2(p; a, q, x, \omega) = p \min\{q, \omega - p\} - ka - \frac{qx}{a}.$$

Deterministic maximization of Π_2 with respect to p, the details of which are left to the reader, yields[6]

$$p^*(q, \omega) = \max\{\omega - q, \omega/2\}.$$

We can see that the firm prices more aggressively, once demand at the

[6] Note that the *sunk cost* portions $-ka$ and $-qx/a$ are *irrelevant* for the pricing decision.

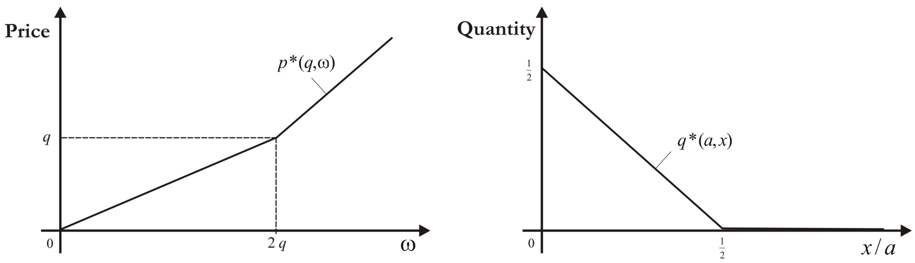

Figure 2: Optimal pricing and production decisions in Example 2 under the assumption that both risks are independent and uniformly distributed.

optimal price reaches the capacity limit (i.e., as soon as $\omega > 2q$), cf. Figure 2. Using this result we can write the firm's expected profits at time $t = 1$,

$$\begin{aligned}\bar{\Pi}_1(q; a, x) &= E\Pi_2(p^*(q, \tilde{\omega}); a, q, x, \tilde{\omega}) \\ &= \int_0^{2q} \left(\frac{\omega}{2} - \frac{x}{a}\right) \frac{\omega}{2} dF(\omega) + \int_{2q}^{\infty} \left(\omega - q - \frac{x}{a}\right) q \, dF(\omega) - ka.\end{aligned}$$

The first-order necessary optimality condition with respect to q is given by

$$2q + \frac{x}{a} = \frac{\int_{2q}^{\infty} \omega dF(\omega)}{1 - F(2q)},$$

an expression that yields candidates for the optimal production quantity $q^*(a, x)$. At $t = 0$, the firm's expected ex-ante profits are given by

$$\bar{\Pi}_0(a) = E\Pi_1(q^*(a, \tilde{x}); a, \tilde{x}).$$

Maximizing the last expression determines the firm's entry and investment decision as a function of the capital intensity k. Note that the firm's investment action a has an informational component, since not only does it increase productive efficiency, it also decreases the technology risk.[7] □

[7]For the special case in which the technology and market risks are (independently) uniformly distributed on $[0, 1]$, i.e., $\tilde{x}, \tilde{\omega} \sim U[0, 1]$, we obtain $q^*(a, x) = [1/2 - x/a]_+$, cf. Figure 2. The firm's optimal action $a^*(k)$ at $t = 0$ is then

$$a^*(k) = \begin{cases} \sqrt{k\left(1 + \sqrt{1 - 64k}\right)}/(4k), & \text{if } k \leq 1/64, \\ 0, & \text{otherwise.} \end{cases}$$

$a^*(k)$ and the resulting investment $ka^*(k)$ are depicted in Figure 3 as a function of the capital intensity k. For $k > \bar{k} = 1/64$ the firm does not enter the market.

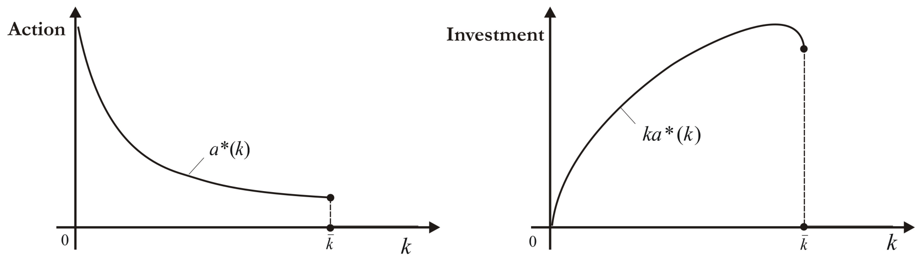

Figure 3: Optimal action $a^*(k)$ and resulting investment $ka^*(k)$ in Example 2 as a function of k under the assumption that both risks are independent and uniformly distributed.

3 Axioms of Choice

So far we have implicitly assumed that actions can be selected by evaluating an expected utility function. We have seen that if the decision maker chooses an action $a \in \mathcal{A}$, then in fact she singles out a distribution of consequences, which we will refer to as a *lottery*. The decision maker thus prefers an action a over another action $\hat{a}$ (where $a, \hat{a} \in \mathcal{A}$), if and only if she prefers the probability distribution implied by a over the one implied by $\hat{a}$. Let $\mathcal{P}$ represent the space of all lotteries over outcomes in a given set $\mathcal{C}$. We will now examine conditions under which preferences over elements of $\mathcal{P}$ can actually be represented by expected utilities.

Definition 1 *A* preference preordering $\preceq$ *over elements of a set $\mathcal{S}$ is a binary relation, so that (i) $s \preceq s$ for any $s \in \mathcal{S}$ (reflexivity); (ii) For all $s_1, s_2, s_3 \in \mathcal{S}$: $s_1 \preceq s_2$ and $s_2 \preceq s_3$ implies that $s_1 \preceq s_3$ (transitivity).*

If for two actions $s_1, s_2 \in \mathcal{S}$ it is $s_1 \preceq s_2$ and $s_2 \preceq s_1$, then we write $s_1 \sim s_2$ denoting *indifference* between s_1 and s_2. If $s_1 \sim s_2$ implies that $s_1 = s_2$ then we call $\preceq$ an *ordering* (in contrast to a *pre*ordering, for which indifference $\sim$ only defines an equivalence class of elements). If $s_1 \preceq s_2$ but not $s_2 \preceq s_1$, then we denote the implied *strict preference* of s_2 over s_1 by $s_1 \prec s_2$. Preferences *per se* already imply the existence of utility functions in many cases.[8]

[8]Note the strict preference $\prec$ does not represent a preference preordering according to Definition 1 as it is not reflexive. It is of course possible to build all the theory using $\prec$ instead of $\preceq$ as the basic binary preference relation, cf. Fishburn (1970) and Kreps (1988).

Proposition 1 *For any set $\mathcal{S}$ the following two statements are equivalent:*[9]

(i) $\preceq$ *is a preference preordering of* $\mathcal{S}$ *and there is a countable subset* $\mathcal{Z} \subset \mathcal{S}$ *such that: for any* $s, \hat{s} \in \mathcal{S}$ *with* $s \prec \hat{s}$ *there exists* $z \in \mathcal{Z}$ *such that* $s \preceq z \preceq \hat{s}$.

(ii) There exists a function $u : \mathcal{S} \to \mathbb{R}$ *such that for any* $s, \hat{s} \in \mathcal{S}$:

$$s \preceq \hat{s} \quad \Leftrightarrow \quad u(s) \leq u(\hat{s}). \tag{4}$$

Proof. (i)$\Rightarrow$(ii): Let $\mathcal{Z} = \{z_1, z_2, \ldots\} \subset \mathcal{S}$ be $\preceq$-dense in $\mathcal{S}$. For each $z_k \in \mathcal{Z}$ ($k \in \{1, 2, \ldots\}$) let $\delta(z_k) = 2^{-k}$ and for any $s \in \mathcal{S}$ let $\bar{\mathcal{Z}}(s) = \{z \in \mathcal{Z} : s \preceq z\}$ and $\underline{\mathcal{Z}}(s) = \{z \in \mathcal{Z} : z \preceq s\}$. By the transitivity of the preference relation,

$$s \preceq \hat{s} \quad \Rightarrow \quad \underline{\mathcal{Z}}(s) \subseteq \underline{\mathcal{Z}}(\hat{s}) \text{ and } \bar{\mathcal{Z}}(\hat{s}) \subseteq \bar{\mathcal{Z}}(s). \tag{5}$$

Let us now introduce the function

$$u(s) = \sum_{z \in \underline{\mathcal{Z}}(s)} \delta(z) - \sum_{z \in \bar{\mathcal{Z}}(s)} \delta(z)$$

with values in $[0, 1]$ (by definition of δ). Then, using (5),

$$\begin{aligned} u(\hat{s}) - u(s) &= \sum_{z \in \underline{\mathcal{Z}}(\hat{s})} \delta(z) - \sum_{z \in \bar{\mathcal{Z}}(\hat{s})} \delta(z) - \sum_{z \in \underline{\mathcal{Z}}(s)} \delta(z) + \sum_{z \in \bar{\mathcal{Z}}(s)} \delta(z) \\ &= \begin{cases} +\sum_{z \in \underline{\mathcal{Z}}(\hat{s}) \setminus \underline{\mathcal{Z}}(s)} \delta(z) + \sum_{z \in \bar{\mathcal{Z}}(s) \setminus \bar{\mathcal{Z}}(\hat{s})} \delta(z) \geq 0, & \text{if } s \preceq \hat{s}, \\ -\sum_{z \in \underline{\mathcal{Z}}(s) \setminus \underline{\mathcal{Z}}(\hat{s})} \delta(z) - \sum_{z \in \bar{\mathcal{Z}}(\hat{s}) \setminus \bar{\mathcal{Z}}(s)} \delta(z) \leq 0, & \text{if } \hat{s} \preceq s, \end{cases} \end{aligned}$$

which implies the utility representation (4) of the preference preordering $\preceq$.

(ii)$\Rightarrow$(i): Let $u : \mathcal{S} \to \mathbb{R}$ be a function such that (4) holds for any $s, \hat{s} \in \mathcal{S}$. It is clear that reflexivity holds, since $s \preceq s$ (i.e., $u(s) = u(s)$) for all $s \in \mathcal{S}$. Moreover, transitivity is satisfied, since $s_1 \preceq s_2$ (i.e., $u(s_1) \leq u(s_2)$) and $s_2 \preceq s_3$ (i.e., $u(s_2) \leq u(s_3)$) implies that $s_1 \preceq s_3$ (i.e., $u(s_1) \leq u(s_3)$. Thus, by Definition 1 relation $\preceq$ is indeed a preference preordering. We now need to show that the representation (4) also implies the existence of a

[9] If for a preference preordering $\preceq$ of $\mathcal{S}$ condition (i) is satisfied for some countable subset $\mathcal{Z} \subset \mathcal{S}$, then we say that $\mathcal{S}$ is $\preceq$-*separable*. Correspondingly, the set $\mathcal{Z}$ is said to be $\preceq$-*dense* in $\mathcal{S}$.

$\preceq$-dense subset $\mathcal{Z}$ of $\mathcal{S}$. Let us denote by $\mathcal{Q}$ the countable collection of real intervals with rational endpoints, i.e.,

$$\mathcal{Q} = \{[a, b] : a < b \text{ and } a, b \in \mathbb{Q}\}.$$

For each interval $\mathcal{I} \in \mathcal{Q}$ with $u(\mathcal{S}) \cap \mathcal{I} \neq \emptyset$ select one element $s \in \mathcal{S}$ such that $u(s) \in \mathcal{I}$. Let $\mathcal{F}$ denote the set of all such elements. Clearly, $\mathcal{F}$ is countable. Furthermore, we set

$$\mathcal{R} = \{(c, d) \in \mathcal{S}^2 \setminus \mathcal{F}^2 : c \prec d \text{ and } \nexists\, s \in \mathcal{F} \text{ such that } c \preceq s \preceq d\}.$$

Hence, if $(c, d) \in \mathcal{R}$, then there is no $s \in \mathcal{S}$ that lies in preference strictly between c and d, i.e., for which $c \prec s \prec d$. Otherwise there would be an element $f \in \mathcal{F}$ with $c \prec f \prec d$, since for any point in the open interval $(u(c), u(d))$ there exists by construction an interval $\mathcal{I}$ in $\mathcal{Q}$ with $\mathcal{I} \subset (u(c), u(d))$ which contains that point. As a result, there cannot be any overlap between any intervals $(u(c), u(d))$ with $(c, d) \in \mathcal{R}$, which implies that $\mathcal{R}$ is countable. Hence, the set

$$\mathcal{G} = \{s \in \mathcal{S} \setminus \mathcal{F} : \exists\, \hat{s} \in \mathcal{S} \text{ such that } (s, \hat{s}) \in \mathcal{R} \text{ or } (\hat{s}, s) \in \mathcal{R}\},$$

is countable, which implies that the union $\mathcal{Z} = \mathcal{F} \cup \mathcal{G}$ is also countable. Now, if $s, \hat{s} \in \mathcal{S} \setminus \mathcal{Z}$ and $s \prec \hat{s}$, then there exists an element $z \in \mathcal{Z}$ such that $s \preceq z \preceq \hat{s}$. In other words, the set $\mathcal{Z} \subset \mathcal{S}$ is $\preceq$-dense in $\mathcal{S}$. ■

When the set $\mathcal{S}$ is countable, then condition (i) in Proposition 1 becomes trivial by setting $\mathcal{Z} = \mathcal{S}$. The following classical example illustrates the impossibility of finding a utility representation for the preferences when $\mathcal{S}$ is not $\preceq$-separable.

Example 3 Consider a decision maker with *lexicographic preferences* $\preceq$, defined on the set $\mathcal{S} = [0, 1] \times [0, 1]$, such that

$$(s_1, s_2) \preceq (\hat{s}_1, \hat{s}_2) \quad \overset{\text{def}}{\Leftrightarrow} \quad s_1 \leq \hat{s}_1 \text{ or } (s_1 = \hat{s}_1 \text{ and } s_2 \leq \hat{s}_2).$$

The decision maker thus prefers an improvement in s_1 more than *any* improvement in s_2. Let us for a moment assume that there exists a utility function $u : \mathcal{S} \to \mathbb{R}$ that represents these preferences according to (4). We now show that this inevitably leads to a contradiction. Note first that $(s_1, 0) \prec (s_1, 1)$ and therefore $u((s_1, 0)) < u((s_1, 1))$ for all $s_1 \in [0, 1]$. If we let

$$\Delta(s_1) = u((s_1, 1)) - u((s_1, 0)),$$

then $\Delta(s_1) > 0$ for all $s_1 \in [0,1]$. As a result, the range $[0,1]$ of the first coordinate can be written as a union of the subsets $\mathcal{S}_{1k} = \{s_1 : \Delta(s_1) \geq 1/k\}$,

$$[0,1] = \bigcup_{k=1}^{\infty} \mathcal{S}_{1k}.$$

Since the interval $[0,1]$ is uncountable, some of the sets $\mathcal{S}_{1k}$ have to be uncountable as well. Let $\mathcal{S}_{1\bar{k}}$ be such an uncountable set for an appropriate $\bar{k} \in \{1,2,\ldots\}$. Let $\bar{\Delta} = u((1,1)) - u((0,0))$ be the largest possible utility difference between any two elements in $\mathcal{S}$ and let $K > \bar{k}\bar{\Delta} + 1$ be an integer. Then for any K elements $\sigma_1, \ldots, \sigma_K \in \mathcal{S}_{1\bar{k}}$ with $\sigma_1 < \sigma_2 < \cdots < \sigma_K$ we have

$$u((\sigma_k, 0)) - u((\sigma_{k-1}, 0)) > u((\sigma_{k-1}, 1)) - u((\sigma_{k-1}, 0)) > 1/\bar{k}$$

for any $k \in \{2, \ldots, K\}$, so that

$$\begin{aligned} \bar{\Delta} &= u((1,1)) - u((0,0)) \\ &= [u((1,1)) - u((\sigma_K, 0))] + [u((\sigma_K, 0)) - u((\sigma_{K-1}, 0))] + \cdots \\ &\quad + [u((\sigma_2, 0)) - u((\sigma_1, 0))] + [u((\sigma_1, 0)) - u((0,0))] \\ &> 0 + 1/\bar{k} + \cdots + 1/\bar{k} + 0 = (K-1)/\bar{k} > \bar{\Delta}, \end{aligned}$$

i.e., a contradiction. A utility representation of lexicographic preferences is therefore not possible. The intuition is that because of the nonseparability of the choice set with respect to $\preceq$, any finite difference in the first attribute must yield an unbounded utility difference, which results from adding up the uncountably many finite utility differences (generated by variations in s_2 for each fixed s_1) in between. $\square$

It turns out that the utility representation (4) of a preference preordering $\preceq$ is unique up to a strictly increasing transformation.

Proposition 2 *Let $\mathcal{S}$ be a set and $u, \hat{u}$ two utility representations of a preference preordering $\preceq$ of $\mathcal{S}$ in the sense of (4). Then there is a function $\varphi : \mathbb{R} \to \mathbb{R}$ such that (i) φ is increasing on $\{v : v = u(s)$ for some $s \in \mathcal{S}\}$. (ii) $\hat{u} = \varphi \circ u$. — In addition, for any increasing function $\phi : \mathbb{R} \to \mathbb{R}$, the function $\phi \circ u$ is a representation of $\preceq$ in the sense of (4).*

Proof. (ii) Let $\mathcal{V} = \{v : v = u(s)$ for some $s \in \mathcal{S}\}$ and define a function $\varphi : \mathbb{R} \to \mathbb{R}$ such that for any $v \in \mathcal{V}$: $\varphi(v) = \hat{u}(s)$, where s is such that

$v = u(s)$. Moreover, on $\mathbb{R} \setminus \mathcal{V}$ define φ such that it is monotonic (which is possible as a consequence of the separability implied by Proposition 1). (i) Let $v_1, v_2 \in \mathcal{V}$ such that $v_1 \leq v_2$. Then by virtue of the fact that $u, \hat{u}$ are utility representations of the form (4) over $\mathcal{S}$ it is

$$v_1 \leq v_2 \;\Leftrightarrow\; s_1 \leq s_2 \;\Leftrightarrow\; \varphi(u(s_1)) = \hat{u}(s_1) \leq \varphi(u(s_2)) = \hat{u}(s_2),$$

where $v_i = u(s_i)$ for $i = 1, 2$. In addition, for any increasing function $\phi : \mathbb{R} \to \mathbb{R}$ we have that

$$s_1 \leq s_2 \;\Leftrightarrow\; u(s_1) \leq u(s_2) \;\Leftrightarrow\; \phi(u(s_1)) \leq \phi(u(s_2)),$$

which completes the proof. ■

In what follows we present one approach to the ordering of the actions based on objective probabilities by Von Neumann and Morgenstern (1944).[10] The axiomatic approach to choice yields a *representation of preferences.* We will show that if a decision maker's preferences satisfy three basic axioms (completeness, continuity, independence), then it is possible to represent her preferences in an expected utility form. For modelling practice, the choice-theoretic axiomatizations are likely to be only of limited relevance. We are interested in these approaches only insofar as they allow us a representation of preferences over actions, which is – as pointed out earlier – implied by a representation of preferences over lotteries.

To avoid mathematical complexities it is convenient to assume that the state space Ω and the action space $\mathcal{A}$ be *finite*.[11] Let the space of consequences be of the form $\mathcal{C} = \mathcal{A} \times \mathcal{X} \times \Omega$, where the outcome space $\mathcal{X} = \{x_1, \ldots, x_N\}$ is finite containing N elements. For each action-state pair $(a, \omega) \in \mathcal{A} \times \Omega$, the outcome $\tilde{x}$ is a random variable with support $\mathcal{X}$. The set of all probability measures over the outcomes can be represented by the $(N-1)$-dimensional simplex $\Delta \subset \mathbb{R}^N_+$. Since there are $|\Omega|$ states, the agent's

[10] A different approach using subjective probabilities is due to Savage (1954). Anscombe and Aumann (1963) derived subjective probabilities based on a set of axioms that is very similar to those by Von Neumann and Morgenstern.

[11] Kreps (1988, pp. 52–63) discusses the Von Neumann-Morgenstern approach for measurable utilities, when the set of lotteries is infinite. The underlying theory, which involves the notion of "mixture spaces," is due to Herstein and Milnor (1953). Kreps (1988, pp. 99–111) also treats the Anscombe-Aumann approach (termed "horse race lotteries and roulette wheels") in the same mixture-space framework. A standard reference in choice theory is Fishburn (1970).

preferences over actions correspond to preferences over lotteries in the $|\Omega|$-fold cartesian product $\Delta^{|\Omega|} = \mathcal{P} \subset \mathbb{R}^{|\Omega|\,N}$. Each element $p \in \mathcal{P}$ is thereby of the form $p = (p_1^\omega, \ldots, p_N^\omega)_{\omega \in \Omega}$, where for all $\omega \in \Omega$:[12]

$$\sum_{k=1}^{N} p_k^\omega = 1.$$

To be very clear, choosing an action a in this framework corresponds to selecting an element p of $\mathcal{P}$, which specifies a distribution of payoffs for each possible state $\omega \in \Omega$.

Von Neumann and Morgenstern (1944, pp. 24–29, 619–632) showed that the following three axioms on a preference preordering over the space of lotteries $\mathcal{P}$ (with outcomes in $\mathcal{X}$) are sufficient for a representation of the expected utility in the form (1).

Axiom 1 (Completeness) *The preference preordering $\preceq$ of $\mathcal{P}$ is complete, i.e., for any $p, q \in \mathcal{P}$ either $p \preceq q$ or $q \preceq p$ (or both).*

Axiom 2 (Continuity) *The preference preordering of $\mathcal{P}$ is continuous, i.e., for any $p, q, r \in \mathcal{P}$ such that $p \preceq q \preceq r$ there exists $t \in [0, 1]$ such that $tp + (1 - t)r \sim q$.*

Axiom 3 (Independence) *The preference preordering is independent of irrelevant alternatives, i.e., for any $p, q, r \in \mathcal{P}$ and $t \in (0, 1)$: $p \prec q \Rightarrow tp + (1 - t)r \prec tq + (1 - t)r$.*

Given these three axioms we are now ready to formulate the main representation theorem. We present here a rather general version (though, only for finite choice sets) that includes state-dependent utilities.

Proposition 3 *Let $\preceq$ be a complete preference preordering of $\mathcal{P}$ satisfying the continuity and independence axioms. (i) Then $\preceq$ can be represented in expected utility form, i.e., there is a function $u : \mathcal{X} \times \Omega \to \mathbb{R}$ such that for*

[12] The element $p \in \mathcal{P}$ can be represented in the form of a *Markov matrix*: instead of writing p as a column vector, each element in $\mathcal{P}$ can be expressed as an $(N \times |\Omega|)$-matrix $\mathbf{P} = [p_k^\omega]_{(k,\omega)}$, whose elements are probabilities and whose columns sum to one.

any $p, q \in \mathcal{P}$:[13]

$$p \preceq q \quad \Leftrightarrow \quad \sum_{\omega} \left(\sum_{k} p_k^{\omega} u(x_k, \omega) \right) \leq \sum_{\omega} \left(\sum_{k} q_k^{\omega} u(x_k, \omega) \right). \tag{6}$$

(ii) The function u in part (i) is unique up to a positive linear transformation: For any u, v which represent $\preceq$ in the expected utility there exist constants α, β with $\alpha > 0$, such that $u = \alpha v + \beta$.

Proof. (i) The proof of the first part proceeds in six steps. Assume for convenience that in $\mathcal{P}$ there exists a *best* lottery $\bar{p}$ and a *worst* lottery $\underline{p}$, such that $\underline{p} \prec \bar{p}$ and for any $p \in \mathcal{P}$: $\underline{p} \preceq p \preceq \bar{p}$. (Should $\underline{p} \sim \bar{p}$, then $p \sim q$ for any $p, q \in \mathcal{P}$ and the expected utility representation (6) holds trivially for $u = 0$.)

Step 1: If $p \prec q$ and $t \in (0, 1)$, then $p \prec tp + (1-t)q \prec q$.

The first claim means that a strict mixture of two lotteries $p, q \in \mathcal{P}$ achieves an intermediate preference. This follows from the independence axiom: indeed, for $r = p$ we obtain from Axiom 3: $p \prec tp + (1-t)q$ for all $t \in (0, 1)$. Similarly, for $r = q$ independence yields $tp + (1-t)q \prec q$, which concludes the first step.

Step 2: For any $s, t \in (0, 1)$: $s\bar{p} + (1-s)\underline{p} \prec t\bar{p} + (1-t)\underline{p} \;\Leftrightarrow\; s < t$.

$\Leftarrow$: Let $s < t$. Then for $\theta = (t-s)/(1-s) \in (0, 1)$ we have that $t\bar{p} + (1-t)\underline{p} = \theta\bar{p} + (1-\theta)(s\bar{p} + (1-s)\underline{p})$, since $\theta + (1-\theta)s = t$ and $(1-\theta)(1-s) = 1-t$. By Step 1, it is furthermore $r = s\bar{p} + (1-s)\underline{p} \prec \bar{p}$, so that by applying Step 1 again for $r \prec \bar{p}$: $r \prec \theta\bar{p} + (1-\theta)r = t\bar{p} + (1-t)\underline{p}$, which is nothing else than $(1-s)\underline{p} \prec t\bar{p} + (1-t)\underline{p}$. $\Rightarrow$: Assume that the converse is true, i.e., $s \geq t$. If $s = t$, then naturally $s\bar{p} + (1-s)\underline{p} \sim t\bar{p} + (1-t)\underline{p}$, which is a contradiction. Similarly, if $s > t$, then we can show as before that $t\bar{p} + (1-t)\underline{p} \prec s\bar{p} + (1-s)\underline{p}$, which is also a contradiction. — This proves

[13] Given the probability space $(\Omega, P(\Omega), \mu)$, the representation (6) is equivalent to the standard expected utility representation,

$$p \preceq q \quad \Leftrightarrow \quad \sum_{\omega} \mu(\omega) \left(\sum_{k} p_k^{\omega} \hat{u}(x_k, \omega) \right) \leq \sum_{\omega} \mu(\omega) \left(\sum_{k} q_k^{\omega} \hat{u}(x_k, \omega) \right),$$

as long as one chooses $\hat{u} : \mathcal{X} \times \Omega \to \mathbb{R}$ such that $u(x_k, \omega) = \mu(\omega)\hat{u}(x_k, \omega)$.

our second claim.

Step 3: For any $p \in \mathcal{P}$ there is a unique $\hat{t}(p) \in [0,1]$ such that $\hat{t}\bar{p}+(1-\hat{t})\underline{p} \sim p$.

Existence follows directly from Axiom 2 and the fact that by assumption $\underline{p} \preceq p \preceq \bar{p}$ for all $p \in \mathcal{P}$. Uniqueness can be obtained as follows. Assume that $\hat{s}, \hat{t} \in (0,1)$ are such that $\hat{s}\bar{p} + (1-\hat{t})\underline{p} \sim p \sim \hat{t}\bar{p} + (1-\hat{t})\underline{p}$. (Note that for $\hat{t} \in \{0,1\}$ or $\hat{s} \in \{0,1\}$ uniqueness holds trivially.) Without loss of generality we can assume that $\tilde{s} \leq \tilde{t}$. If $\tilde{s} < \tilde{t}$ then by Step 2 we have that $\hat{s}\bar{p} + (1-\hat{s})\underline{p} \prec \hat{t}\bar{p} + (1-\hat{t})\underline{p}$, a contradiction. Thus $\hat{s} = \hat{t}$ which implies uniqueness.

Step 4: The function $\Phi : \mathcal{P} \to \mathbb{R}$ with $\Phi(p) = \hat{t}(p)$ represents the preference preordering $\preceq$ of $\mathcal{P}$, i.e., relation (4) in Proposition 1 holds for $\mathcal{S} = \mathcal{P}$.

Indeed for any $p, q \in \mathcal{P}$ we have: $p \preceq q$ if and only if $\hat{t}(p)\bar{p} + (1-\hat{t}(p))\underline{p} \preceq \hat{t}(q)\bar{p} + (1-\hat{t}(q))\underline{p}$. Step 2 therefore implies that

$$p \preceq q \quad \Leftrightarrow \quad \hat{t}(p) \leq \hat{t}(q),$$

which proves our claim.

Step 5: The expected utility function Φ is linear on $\mathcal{P}$.

Let $p, q \in \mathcal{P}$ and $\gamma \in [0,1]$. For our claim to hold it is enough to show that $\Phi(\gamma p + (1-\gamma)q) = \gamma\Phi(p) + (1-\gamma)\Phi(q)$. By construction of the function Φ it is

$$\begin{aligned} p &\sim \Phi(p)\bar{p} + (1-\Phi(p))\underline{p}, \\ q &\sim \Phi(q)\bar{p} + (1-\Phi(q))\underline{p}, \end{aligned}$$

and thus by using Axiom 3 (independence):

$$\begin{aligned} \gamma p + (1-\gamma)q &\sim \gamma\left(\Phi(p)\bar{p} + (1-\Phi(p))\underline{p}\right) + (1-\gamma)\left(\Phi(q)\bar{p} + (1-\Phi(q))\underline{p}\right) \\ &\sim \left[\gamma\Phi(p) + (1-\gamma)\Phi(q)\right]\bar{p} + \left[1 - (\gamma\Phi(p) + (1-\gamma)\Phi(q))\right]\underline{p}, \end{aligned}$$

so that necessarily $\Phi(\gamma p + (1-\gamma)q) = \gamma\Phi(p) + (1-\gamma)\Phi(q)$, which establishes the linearity of Φ.

Step 6: The expected utility function $\Phi(\cdot)$ has the representation in (6).

Note first that as a direct consequence of the linearity of Φ for any $p \in \mathcal{P}$ we can write $\Phi(p)$ as a linear combination of the elements of p:

$$\Phi(p) = \Phi\left(\sum_{\omega,k} p_k^\omega \rho_k^\omega\right),$$

where ρ_k^ω are appropriate coefficients. Let $u : \mathcal{X} \times \Omega \to \mathbb{R}$ be a map that assigns $u(x_k, \omega) = \rho_k^\omega$. The function u can be interpreted as a *state-dependent* utility function. Thus, we obtain

$$\Phi(p) = \sum_\omega \left(\sum_k p_k^\omega u(x_k, \omega)\right),$$

which completes the last step.

(ii) Let us assume, essentially for the same reasons as in part (i), that there exists a *best* lottery $\bar{p}$ and a *worst* lottery $\underline{p}$, such that $\underline{p} \prec \bar{p}$ and for any $p \in \mathcal{P}$: $\underline{p} \preceq p \preceq \bar{p}$. From part (i) it is clear that if u is a representation of $\preceq$, then (6) holds. Let us first show that (6) then also holds for $v = \alpha u + \beta$, where α, β are arbitrary real constants with $\alpha > 0$. Indeed,

$$\begin{aligned}
\sum_\omega \left(\sum_k p_k^\omega v(x_k, \omega)\right) &= \alpha \sum_\omega \left(\sum_k p_k^\omega u(x_k, \omega)\right) + \beta|\Omega| \\
&\leq \alpha \sum_\omega \left(\sum_k q_k^\omega u(x_k, \omega)\right) + \beta|\Omega| \\
&= \sum_\omega \left(\sum_k q_k^\omega v(x_k, \omega)\right),
\end{aligned}$$

if and only if the inequality in (6) holds. If on the other hand both u and v are chosen such that the expected utility representation (6) for $\preceq$ is valid, then one can find constants β and $\alpha > 0$ such that $u = \alpha v + \beta$. Indeed, given any $p \in \mathcal{P}$ we can define $\vartheta(p) = (\Phi(p) - \Phi(\underline{p}))/(\Phi(\bar{p}) - \Phi(\underline{p})) \in [0, 1]$. Since $\Phi(p)$ is a representation of the preference preordering $\preceq$ of $\mathcal{P}$, it is by part (i) linear. Thus

$$\vartheta(p)\Phi(\bar{p}) + (1 - \vartheta(p))\Phi(\underline{p}) = \Phi(\vartheta(p)\bar{p} + (1 - \vartheta(p))\underline{p})$$

and necessarily $p \sim \vartheta(p)\bar{p} + (1 - \vartheta(p))\underline{p}$. Similarly, since v is part of a representation of $\preceq$ in the form (6) with expected utility $\Psi(\cdot)$ defined on $\mathcal{P}$, we must have

$$\begin{aligned} \Psi(p) &= \Psi(\vartheta(p)\bar{p} + (1 - \vartheta(p))\underline{p}) \\ &= \vartheta(p)\Psi(\bar{p}) + (1 - \vartheta(p))\Psi(\underline{p}) \\ &= \vartheta(p)\left(\Psi(\bar{p}) - \Psi(\underline{p})\right) + \Psi(\underline{p}), \end{aligned} \tag{7}$$

so that $\Psi(p) = \alpha\Phi(p) + \gamma$, with $\alpha = (\Psi(\bar{p}) - \Psi(\underline{p}))/(\Phi(\bar{p}) - \Phi(\underline{p}))$ and $\gamma = \Psi(\underline{p}) - \Phi(\underline{p})\alpha$. Hence part (ii) of our proposition obtains by setting $\beta = \gamma/|\Omega|$. ■

Remark Even though probably quite clear by now, it may still be useful to clarify the relation of Proposition 3 to a representation of preferences over the decision maker's *actions*. Let $\Phi : \mathcal{P} \to \mathbb{R}$ be an expected utility representation, i.e., for any $p \in \mathcal{P}$:

$$\Phi(p) = \sum_{\omega} \left(\sum_{k} p_k^{\omega} u(x_k, \omega) \right)$$

for an appropriate state-dependent $u : \mathcal{X} \times \Omega \to \mathbb{R}$. Furthermore, let $\psi : \mathcal{A} \to \mathcal{P}$ be a mapping from the decision maker's action space to the space of lotteries. This mapping is normally implied by the setup of a particular problem. Then we can express the expected utility (cf. footnote 13 on page 14) form (6) of any action $a \in \mathcal{A}$ as

$$EU(a) = \Phi(\psi(a)).$$

The expected utility $EU(\cdot)$ is unique up to a positive linear transformation. It represents a complete preference preordering of the action set $\mathcal{A}$.

4 Risk Aversion

Consider an agent with utility function $u : \mathcal{X} \to \mathbb{R}$ and wealth w. To keep the discussion as simple as possible, we restrict attention to uncertain events with *monetary* outcomes in a set $\mathcal{X} \subseteq \mathbb{R}$. We further assume that u is twice differentiable and strictly increasing.[14] When confronted with a money lottery of uncertain payoff $\tilde{x}$, his expected utility is $Eu(w + \tilde{x})$.

[14]Smoothness assumptions on utility functions are quite uncritical as they are hard to refute given any finite amount of data on choice outcomes.

Definition 2 *The agent is* risk averse *if he dislikes any zero-mean lottery, i.e., if for* $\tilde{x}$ *with* $E\tilde{x} = 0$ *it is* $Eu(w + \tilde{x}) < u(w)$ *for all wealth levels* w*.*

Naturally, the agent is *risk loving* if he strictly prefers accepting any zero mean lottery over doing nothing. In case of indifference for all wealth levels the agent is *risk neutral.* We now provide a full characterization of risk aversion.

Proposition 4 *The agent is risk averse if and only if his utility function is strictly concave.*

Proof. Assume first that the agent is risk averse. Then for any zero-mean random variable $\tilde{x}$ and any wealth level w we have that $Eu(w + \tilde{x}) < u(w)$. Let $\tilde{y} = w + \tilde{x}$ and thus $Eu(\tilde{y}) < u(E\tilde{y})$ which by Jensen's inequality (cf. Appendix) implies strict concavity. Similarly, if the agent's utility function u is strictly concave, the converse follows immediately, again using Jensen's inequality. ∎

Clearly, a risk-averse agent is willing to pay a positive amount π in order *not* to be exposed to a *pure* risk $\tilde{x}$ (with $E\tilde{x} = 0$). This amount is called his *risk premium* $\pi = \pi(u; w, \tilde{x})$ and can be computed as follows:

$$Eu(w + \tilde{x}) = u(w - \pi(u; w, \tilde{x})), \tag{8}$$

where w is the agent's initial wealth and u his utility function (cf. Figure 4). If the agent is risk loving, then π might become negative. If instead of pure risk, the agent is exposed to a random payoff $\tilde{y} = \mu + \tilde{x}$, so that $E\tilde{y} = \mu$, then the amount an agent would require to be paid in order to make him indifferent between accepting the money lottery or taking his wealth w for sure is called his *certainty equivalent* $CE = CE(u; w, \tilde{y})$. The certainty equivalent satisfies

$$Eu(w + \tilde{y}) = u(w + CE(u; w, \tilde{y})). \tag{9}$$

Comparing (8) and (9) we find that $w - \pi = w - \mu + CE$ or in other words,

$$CE(u; w, \tilde{y}) = \mu - \pi(u; w + \mu, \tilde{x}).$$

We would like to order utility functions in a way that leads to unambiguous changes in the risk premium, independent of the underlying risk.[15]

[15] Let us consider a simple *numerical example* to see how the algebra works: an individual

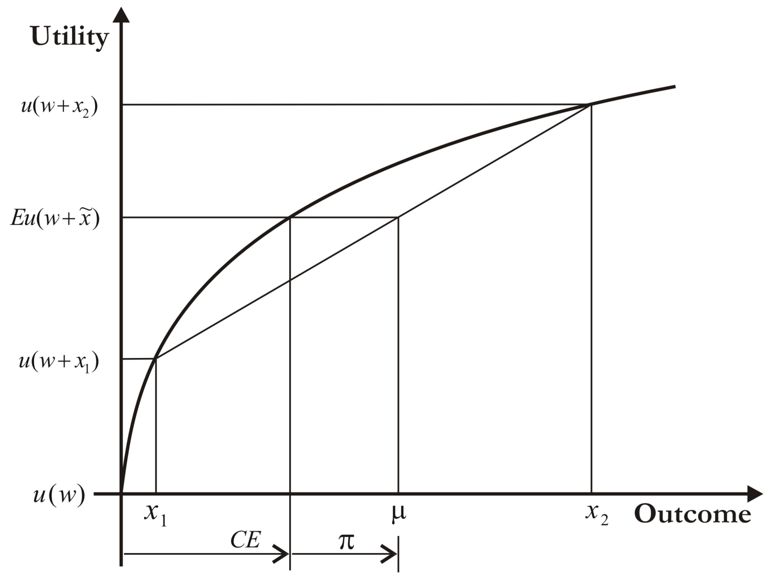

Figure 4: Risk premium π and certainty equivalent CE of a money lottery $\tilde{x} \in \{x_1, x_2\}$ with $E\tilde{x} = \mu$ and initial wealth w.

Definition 3 *Agent U with utility function u is* more risk averse *than agent V with utility function v, if at any equal wealth level agent U rejects all lotteries that V rejects.*

The following statement provides necessary and sufficient conditions for agent U to be more risk averse than agent V. Before we formulate this proposition it is useful to define the *Arrow-Pratt coefficient of absolute risk aversion* for, say, agent U:

$$\rho_A(u; w) = -\frac{u''(w)}{u'(w)}. \tag{10}$$

Proposition 5 *Assume that agents U and V have the same initial wealth w and suppose that their utility functions u and v are twice differentiable. Then the following statements are equivalent:*

---

of wealth $w = 0$ faces a money lottery $\tilde{y}$, which with equal probability of $1/2$ yields $y_1 = 0$ or $y_2 = -16$. Preferences are represented by a utility function $u(y) = \sqrt{16 + y}$ for $y \geq -16$. The expected payoff of the lottery $\tilde{y}$ is $\mu = E\tilde{y} = -8$ and its certainty equivalent can be determined from (9), which gives $CE(u; 0, \tilde{y}) = -12$. The *pure risk* associated with the lottery $\tilde{y}$ is given by $\tilde{x} = \tilde{y} - \mu$. From (8) we obtain the risk premium at wealth μ: $\pi(u; \mu, \tilde{x}) = 4$. Thus, we have verified that indeed $CE(u; w, \tilde{y}) = \mu - \pi(u; w + \mu, \tilde{x})$.

(i) Agent U is more risk averse than agent V.

(ii) There is a strictly increasing concave function φ such that $u = \varphi \circ v$.

(iii) Agent U's absolute risk aversion is larger than agent V's absolute risk aversion, i.e., $\rho_A(u; w) \geq \rho_A(v; w)$ for all w.

(iv) The risk premium that agent U is willing to pay exceeds the risk premium that agent V is willing to pay, i.e., $\pi(u; w) \geq \pi(v; w)$ for all w.

Proof. (i)$\Leftrightarrow$(iv): Agent U is more risk averse than agent V if and only if (by definition) for any random payoff $\tilde{y}$ with $E\tilde{y} = \mu$ and for any wealth level w:

$$v(w+\mu-\pi(v)) = Ev(w+\tilde{y}) < v(w) \Rightarrow u(w+\mu-\pi(u)) = Eu(w+\tilde{y}) < u(w),$$

where $\pi(u) = \pi(u; w)$ and $\pi(v) = \pi(v; w)$ are the risk premiums for agents U and V. This is equivalent with

$$\mu - \pi(v) < 0 \;\Rightarrow\; \mu - \pi(u) < 0,$$

which is equivalent to $\pi(u) \geq \pi(v)$, as claimed. (ii)$\Leftrightarrow$(iii): Since any agent prefers a higher payoff to less (u, v are strictly increasing) and given that their utility functions are twice differentiable, there exists a twice differentiable strictly increasing function $\varphi : \mathbb{R} \to \mathbb{R}$, such that $u = \varphi \circ v$. Taking the first and second derivative of u we obtain therefore $u' = \varphi'(v)v'$ and $u'' = \varphi''(v)(v')^2 + \varphi'(v)v''$. Dividing the expression for u'' by u' and using the expression for u' we obtain

$$\frac{u''}{u'} = \frac{\varphi''(v)}{\varphi'(v)v'} + \frac{v''}{v'},$$

or by introducing the definition of the Arrow-Pratt coefficient of absolute risk aversion,

$$\rho_A(u; w) = \rho_A(v; w) - \frac{\varphi''(v)}{\varphi'(v)v'},$$

so that $\rho_A(u; w) \leq \rho_A(v; w)$ if and only if $\varphi''(v(w)) \leq 0$. (ii)$\Rightarrow$(iv): There is a strictly increasing and concave function $\varphi : \mathbb{R} \to \mathbb{R}$ such that $u = \varphi \circ v$. Thus, for any wealth level w and random payoff $\tilde{y}$ with $E\tilde{y} = \mu$ we have $Eu(w+\tilde{y}) = u(w+\mu-\pi(u)) = \varphi(v(w+\mu-\pi(u))$. Using Jensen's inequality, since φ is concave: $u(w + \mu - \pi(u)) = E\varphi(v(w + \tilde{y})) \leq \varphi(Ev(w + \tilde{y})) =$

$\varphi(v(w+\mu-\pi(v))) = u(w+\mu-\pi(v))$. As a result, $\pi(u) > \pi(v)$. (ii)$\Leftarrow$(iv): By definition of the risk premium, $Ev(w+\tilde{y}) = v(w+\mu-\pi(v))$. Thus $\varphi(Ev(w+\tilde{y})) = u(w+\mu-\pi(v))$. On the other hand, $E\varphi(v(w+\tilde{y})) = u(w+\mu-\pi(u))$. Since by assumption $\pi(u) \geq \pi(v)$, we have that $E\varphi(v(w+\tilde{y})) \leq \varphi(Ev(w+\tilde{y}))$. Thus, by Jensen's inequality (since $\tilde{y}$ and w are arbitrary) φ is a concave function. This concludes the proof of the proposition. ■

Remark The coefficient of *relative* risk aversion is defined as

$$\rho_R(u;w) = w\rho_A(u;w). \tag{11}$$

A family of functions with constant absolute risk aversion (CARA) can be derived by solving the homogenous ordinary differential equation $u'' - \rho u' = 0$ for any real constant ρ. One obtains (up to a positive linear transformation) a representation of CARA functions (with $\rho_A = \rho$):

$$u(w) = -\exp(-\rho w).$$

Similarly one can derive all (CRRA) functions with constant relative risk aversion $\rho_R = \rho$:

$$u(w) = \begin{cases} w^{1-\rho}/(1-\rho), & \text{if } \rho \neq 1, \\ \ln(w), & \text{otherwise.} \end{cases}$$

5 Stochastic Dominance

It is now our goal to order risks in a way that leads to unambiguous changes in the risk premium[16] for a class $\mathcal{U}$ of utility functions u. Given two random variables $\tilde{x}$ and $\tilde{y}$ it is $\pi(u,\tilde{x}) \geq \pi(u,\tilde{y})$ for all $u \in \mathcal{U}$, if and only if

$$Eu(\tilde{x}) \leq Eu(\tilde{y}) \quad \forall u \in \mathcal{U}. \tag{12}$$

Definition 4 *If (12) holds, the risk $\tilde{y}$ is said to* stochastically dominate *$\tilde{x}$ with respect to $\mathcal{U}$, denoted by $\tilde{x} \preceq_{\mathcal{U}} \tilde{y}$. Necessary and sufficient conditions on $\tilde{x}$ and $\tilde{y}$ for (12) to hold are called a* stochastic dominance order (representation) relative to $\mathcal{U}$.

Note that for $\mathcal{U}' \subset \mathcal{U}$ the stochastic dominance order will be weaker, as it has to characterize (12) for less functions. There are two sets of utility functions that are of particular practical interest:

[16] Without loss of generality we assume that wealth w is equal to zero, and we omit it in most of this section.

1. The stochastic dominance order relative to all increasing functions (the set $\mathcal{U}_1$) is called *first-order stochastic dominance.*

2. The stochastic dominance order relative to all increasing and concave functions (the set $\mathcal{U}_2$) is called *second-order stochastic dominance.*

In addition to these two widely used stochastic orders, we will consider *central dominance* in Section 5.3, which proves useful in certain applications. In order to construct stochastic dominance orders, let us first reformulate (12) by representing the set $\mathcal{U}$ in terms of convex combinations of functions from a simpler set, $\mathcal{B} \subset \mathcal{U}$.

Definition 5 *Let $\mathcal{B} = \{b(\cdot, \theta) : \theta \in \Theta\} \subset \mathcal{U}$, where Θ is a compact index set. If for any $u \in \mathcal{U}$ there exists an increasing measurable function $H : \Theta \to \mathbb{R}_+$ and real constants α, β with $\beta > 0$ such that*

$$u(\cdot) = \alpha + \beta \int_\Theta b(\cdot, \theta) dH(\theta), \tag{13}$$

then $\mathcal{B}$ is called a basis *of $\mathcal{U}$. If H is nonnegative and satisfies the normalization condition*

$$\int_\Theta dH(\theta) = 1, \tag{14}$$

it is called the transform of u with respect to $\mathcal{B}$.

Note that given any particular $u \in \mathcal{U}$ if (14) holds, we can interpret H, its transform with respect to $\mathcal{B}$, as a cumulative distribution function of a random variable $\tilde{\theta}$ that assumes values in Θ. Then, relation (13) is equivalent to $u(\cdot) = Eb(\cdot, \tilde{\theta})$. We will now see that using a basis of $\mathcal{U}$ it is possible to considerably simplify (12).

Proposition 6 *If $\mathcal{B}$ is a basis of $\mathcal{U}$, then (12) can be rewritten as*

$$Eb(\tilde{x}, \theta) \leq Eb(\tilde{y}, \theta) \quad \forall \theta \in \Theta. \tag{15}$$

Proof. The proof is almost trivial. Let $\mathcal{B}$ be a basis of $\mathcal{U}$. (12)⇒(15): Since for all $\theta \in \Theta$ we have that $b(\cdot, \theta) \in \mathcal{U}$, relation (12) implies that $Eb(\tilde{x}, \theta) \leq Eb(\tilde{y}, \theta)$ for all $\theta \in \Theta$ (15)⇒(12): Since any $u \in \mathcal{U}$ can be represented as $u(\cdot) = Eb(\cdot, \tilde{\theta})$ for some increasing measurable function $H(\theta)$, relation (12) follows immediately from (15) by convex combination. ■

5.1 First-Order Stochastic Dominance

Let $\mathcal{U}_1 = \{u : x < y \Rightarrow u(x) \leq u(y)\}$ be the set of all increasing utility functions, $u : [a, b] \to \mathbb{R}$ for some real constants $a < b$. We would like to construct a stochastic dominance order $\preceq_1$, so that $\tilde{x} \preceq_1 \tilde{y}$ is equivalent to (12) for $\mathcal{U} = \mathcal{U}_1$. If we let $\Theta = [a, b]$, then the set of step functions

$$\mathcal{B}_1 = \{b(\cdot, \theta) \in \mathcal{U}_1 : b(x, \theta) = \mathbf{1}_{\{x \geq \theta\}}, \theta \in [a, b]\}$$

is a basis of $\mathcal{U}_1$, which can be readily verified. Indeed, for $H(\theta) = u(\theta)$, $\alpha = u(a)$ and $\beta = 1$ we have[17]

$$u(x) = u(a) + \int_a^x du(\theta) = \alpha + \beta \int_a^b b(x, \theta) dH(\theta).$$

By Proposition 6 and linearity of the expectation operator we find that relation (12) with $\mathcal{U} = \mathcal{U}_1$ is equivalent to $E\mathbf{1}_{\{\tilde{x} \geq \theta\}} \leq E\mathbf{1}_{\{\tilde{y} \geq \theta\}}$ for all $\theta \in \Theta$, which in turn is equivalent to

$$F(\theta) \geq G(\theta) \quad \forall \theta \in \Theta, \tag{16}$$

where F is the cdf of $\tilde{x}$ and G the cdf of $\tilde{y}$. Condition (16) thus *characterizes* first-order stochastic dominance: $\tilde{x} \preceq_1 \tilde{y}$ if and only if (16) holds.

MPR and MLR Stochastic Orders. A *sufficient* condition for first-order stochastic dominance is that the *probability ratio* $\pi(\theta) = \text{Prob}(\tilde{x} \leq \theta)/\text{Prob}(\tilde{y} \leq \theta) = F(\theta)/G(\theta)$ is monotonically decreasing in θ. The stochastic order $\preceq_{\text{MPR}}$ is generally referred to as *monotone probability ratio* (MPR) order. Similarly, if the *likelihood ratio* $\ell(\theta) = F'(\theta)/G'(\theta)$ is decreasing in θ, then this is sufficient for first-order dominance. The implied stochastic order $\preceq_{\text{MPR}}$ is the *monotone likelihood ratio* (MLR) order. It turns out (cf. Gollier (2001, pp. 102–104) or Athey (2002)) that $\preceq_{\text{MPR}}$ is weaker than $\preceq_{\text{MLR}}$; in other words: $\tilde{x} \preceq_{\text{MLR}} \tilde{y} \;\Rightarrow\; \tilde{x} \preceq_{\text{MPR}} \tilde{y}$.

5.2 Second-Order Stochastic Dominance

Let $\mathcal{U}_2 = \mathcal{U}_1 \cap \{u : tu(y) + (1-t)u(x) \leq u(ty + (1-t)x), t \in [0, 1]\}$ be the set of all increasing and concave utility functions, $u : [a, b] \to \mathbb{R}$. We would

[17]We assume here for convenience that $u(b) > u(a)$. If $u(b) = u(a)$, then the function $u \in \mathcal{U}_1$ is necessarily constant on $[a, b]$ and relation (13) holds trivially. Note also that any monotone function defined on a compact set is measurable.

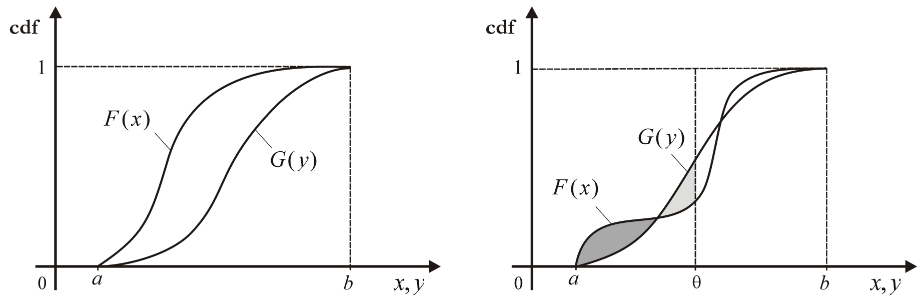

Figure 5: FOSD condition (16) and SOSD condition (18) for comparing two risks $\tilde{x} \sim F$ and $\tilde{y} \sim G$, so that $\tilde{x} \preceq_1 \tilde{y}$ and $\tilde{x} \preceq_2 \tilde{y}$ respectively.

like to construct a stochastic dominance ordering $\preceq_2$, so that $\tilde{x} \preceq_2 \tilde{y}$ is equivalent to (12) for $\mathcal{U} = \mathcal{U}_2$. We first show that

$$\mathcal{B}_2 = \{b(\cdot, \theta) \in \mathcal{U}_2 : b(x, \theta) = \min\{x, \theta\}, \theta \in [a, b]\}$$

is a basis of $\mathcal{U}_2$ by constructing appropriate functions $H(\theta)$. Let α, β be real constants with $\beta > 0$. To get an idea about what requirement appropriate H function needs to satisfy for a given $u \in \mathcal{U}_2$, let us first assume that we have already found such a function. Relation (13) can – using integration by parts – be written as

$$\begin{aligned} u(x) &= \alpha + \beta \int_a^b b(x, \theta) dH(\theta) = \alpha + \beta \int_a^b \min\{x, \theta\} dH(\theta) \\ &= \alpha + \beta \left(xH(b) - aH(a) - \int_a^x H(\theta) d\theta \right), \end{aligned}$$

for all admissible x, so that[18]

$$u'(x) \overset{\text{a.e.}}{=} \beta(H(b) - H(x)).$$

Note that the last relation implies that $u'(b) = 0$, which is not a requirement of being a member of the set $\mathcal{U}_2$. Since the set of twice differentiable concave increasing functions is dense in $\mathcal{U}_2$ it seems reasonable, based on the previous derivations to require that

$$H'(\theta) = -u''(\theta),$$

[18]Note that $u \in \mathcal{U}_2$ is as a convex function differentiable almost everywhere.

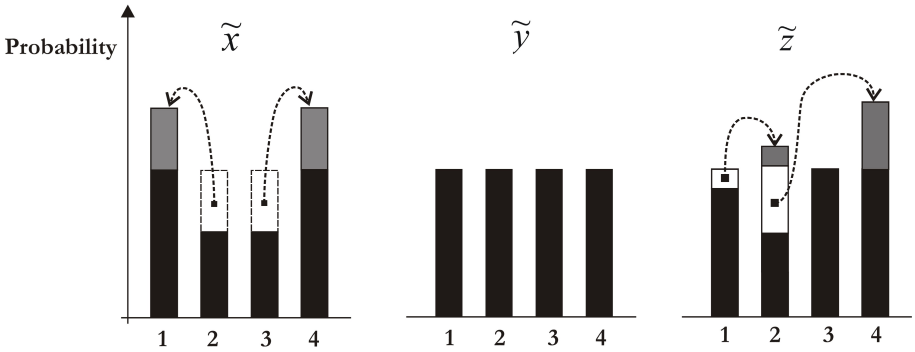

Figure 6: FOSD and SOSD shifts for random variables $\tilde{x}, \tilde{y}, \tilde{z}$ on a discrete support $\{1, 2, 3, 4\}$, so that $\tilde{x} \preceq_2 \tilde{y} \preceq_1 \tilde{z}$. Note that $\tilde{x} \sim_1 \tilde{y}$ and $\tilde{x} \preceq_2 \tilde{z}$.

for such functions (in particular, H should be differentiable then). Integrating the last relation on both sides on $[a, x]$, we obtain that

$$H(x) = u'(a) - u'(x), \tag{17}$$

where we have set $H(a) = 0$ without loss of generality. The reader is urged to verify using (13) that (17) (together with $\alpha = -(au'(a) + u(a))$ and $\beta = 1$) is indeed an appropriate choice of H for *any* $u \in \mathcal{U}_2$. By Proposition 6 relation (12) is therefore equivalent to

$$E\left[\min\{\tilde{x}, \theta\}\right] \leq E\left[\min\{\tilde{y}, \theta\}\right] \quad \forall \theta \in \Theta.$$

In other words,

$$\int_a^\theta F(x)dx \geq \int_a^\theta G(y)dy \quad \forall \theta \in [a, b], \tag{18}$$

where F and G are the cumulative distribution functions for $\tilde{x}$ and $\tilde{y}$ respectively. Condition (18) defines second-order stochastic dominance: $\tilde{x} \preceq_2 \tilde{y}$ if and only if (18) is satisfied.[19] If $E[\tilde{y} - \tilde{x}] = 0$, then a second-order stochastically dominated deterioration of risk is referred to as a *mean-preserving spread*.

Remark Rothschild and Stiglitz (1970) show that a mean-preserving spread

[19]Note that since the identity function $u(x) = x$ is in $\mathcal{U}_2$, the relation $\tilde{x} \preceq_2 \tilde{y}$ means that the expected difference $E[\tilde{y} - \tilde{x}]$ is nonnegative, as a direct consequence of (12).

is equivalent to the addition of independent white noise $\tilde{\varepsilon}$, so that

$$\tilde{x} = \tilde{y} + \tilde{\varepsilon} \tag{19}$$

with $E[\tilde{\varepsilon}|\tilde{y}] = 0$. Using Jensen's inequality and the law of iterated expectations, sufficiency of (19) for (12) can be readily established:

$$Eu(\tilde{x}) = Eu(\tilde{y} + \tilde{\varepsilon}) = E[E[u(\tilde{y} + \tilde{\varepsilon})|\tilde{y}]] \le E[u(\tilde{y} + E[\tilde{\varepsilon}|\tilde{y}])] = Eu(\tilde{y}).$$

Rothschild and Stiglitz also establish necessity in their seminal paper.

First-order stochastic dominance (FOSD) implies second-order stochastic dominance (SOSD), since condition (16) is more restrictive than (18). Figure 5 illustrates the two conditions next to each other. The intuitive difference between FOSD shifts and SOSD shifts can be easily grasped by looking at random variables on a discrete support. Figure 6 provides such an example.

5.3 Central Dominance

Consider the portfolio problem discussed earlier, whereby we assume that the investor is risk averse, so that by Proposition 4 his utility function is strictly concave. The first-order necessary optimality condition for an interior maximum of the expected utility with respect to the investment action a is

$$EU'(a) = E\tilde{r}u'(w + a\tilde{r}) = 0. \tag{20}$$

The interesting question is, what shifts in the distribution of returns would prompt the investor to increase his optimal investment? To describe such shifts let us fix $a \in (0, 1)$ such that (20) holds and let $\tilde{x} = a\tilde{r}$ be the outcome distribution associated with this action.

Definition 6 *The random variable $\tilde{x}$ is* centrally dominated *by the random variable $\tilde{y}$ at w, if*

$$E\tilde{x}u'(w + \tilde{x}) = 0 \;\Rightarrow\; E\tilde{y}u'(w + \tilde{y}) \ge 0, \tag{21}$$

for all $u \in \mathcal{U}_2$.

It is clear that if $\tilde{y}$ centrally dominates $\tilde{x}$ at w, then the investor would increase her investment action $a > 0$ as the distribution of returns shifts from $\tilde{x}/a$ to $\tilde{y}/a$. We will characterize central dominance in the next section using the diffidence theorem (cf. Example 4 on page 29).

6 Comparative Statics under Uncertainty

The term "comparative statics" refers to the practice of determining the impact of parameter changes in models. Analyzing the comparative statics for your model typically yields some of the main insights for your applied research paper. Note that it is often possible to make statements about comparative statics without being able to actually solve the underlying model explicitly.

6.1 The Diffidence Theorem

Consider first the simple case of a model parameter α that can take only two values, $\alpha \in \{\alpha_1, \alpha_2\}$. Let $\varphi(\tilde{x}; \alpha)$ be some function of interest in your model (e.g., a firm's profit) that depends on a random outcome $\tilde{x}$ and the exogenous parameter α. For simplicity, let us write $f(\cdot) = \varphi(\cdot; \alpha_1)$ and $g(\cdot) = \varphi(\cdot; \alpha_2)$ and assume that f and g are measurable functions. We are interested in necessary and sufficient conditions on f, g which guarantee that

$$Ef(\tilde{x}) \geq Eg(\tilde{x}), \tag{22}$$

for any $\tilde{x}$ with support in the compact interval $[a, b]$ (to keep technicalities to a minimum). Since (22) needs to hold for all degenerate random variables $\tilde{x} = x \in [a, b]$ a *necessary* condition is that

$$f(x) \geq g(x), \tag{23}$$

for all $x \in [a, b]$. On the other hand, the last condition is also *sufficient*, since (23) directly implies (22) for any $\tilde{x}$ with support in $[a, b]$. Thus, (22) and (23) are equivalent.

Let us now consider the somewhat more interesting case, where a little more is known about your model. For instance, a firm's expected profit might be zero when $\alpha = \alpha_1$ for *some* randomly distributed outcome, i.e., $Ef(\tilde{x}) = 0$ for a particular $\tilde{x}$ with support in $[a, b]$. With this additional structure in hands, inequality (22) is weakened to

$$\forall \tilde{x} : \quad Ef(\tilde{x}) = 0 \;\Rightarrow\; Eg(\tilde{x}) \leq 0. \tag{24}$$

We can expect that a necessary and sufficient condition for (24) will be weaker than (23), since it only has to hold for random variables $\tilde{x}$ that satisfy the additional zero-expected-profit condition. It turns out that quite a few problems can be rewritten in the form (24) so that a characterization of this relation is of great interest. The following theorem (due to unpublished work by Gollier and Kimball (1997)) provides such a characterization.[20]

Proposition 7 (Diffidence Theorem) *Let $f, g : \mathbb{R} \to \mathbb{R}$ be measurable functions. The following two statements are equivalent. (i) For any real random variable $\tilde{x}$ on a support in $[a, b]$: $Ef(\tilde{x}) = 0 \Rightarrow Eg(\tilde{x}) \leq 0$. (ii) There exists a real constant λ such that $g(x) \leq \lambda f(x)$ for almost all $x \in [a, b]$.*

Proof. The proof given here is somewhat informal: cf. Gollier (2001, pp. 82–86) for details. Let us first assume that there is a pdf $h(\cdot)$ of $\tilde{x}$, defined on $[a, b]$. By definition, $h \geq 0$ and

$$\int_a^b h(x)dx = 1. \tag{25}$$

The condition $Ef(\tilde{x}) = 0$ can be written as

$$\int_a^b f(x)h(x)dx = 0. \tag{26}$$

Thus for any given f, g relation (24) holds if and only if

$$\sup_{h(\cdot)} \int_a^b (g(x) - \lambda f(x) - \mu)\, h(x)dx + \mu \leq 0 \tag{27}$$

for some constants λ and μ, subject to (25)–(26). The last inequality has been obtained from the condition $Eg(\tilde{x}) \leq 0$ by adding the constraints (25)–(26) multiplied by μ and λ respectively. Given that $h \geq 0$ and (25) holds, a *sufficient* condition for (27) is

$$g(x) - \lambda f(x) \leq 0$$

for almost all $x \in [a, b]$ and some λ (which needs to be nonnegative, since $f, g \geq 0$). The last condition is also *necessary*, for if there is no λ such that $g(x) \leq \lambda f(x)$ on $[a, b]$ (almost everywhere), then the LHS in (27) becomes

[20] This theorem is somewhat related to hyperplane separation theorems in Hilbert spaces. For details on such separation theorems, see for instance Aubin (1998, pp. 27–34), Berge (1963, pp. 154–157, pp. 162–167) or more specifically Jewitt (1986). Its peculiar name is related to the concept of "diffidence" which we introduce on page 29.

positive even though the random variable $\tilde{x}$ satisfies (26). ■

An agent of initial wealth w and utility function u dislikes any zero-mean lottery if and only if

$$E\tilde{x} = 0 \;\Rightarrow\; Eu(w + \tilde{x}) \leq u(w). \tag{28}$$

If the last condition holds for all wealth levels w, then by Definition 2 we obtain (weak) risk aversion, which by Proposition 4 is equivalent to u being concave.[21] If (28) only holds for certain wealth levels, then concavity of u (i.e., risk aversion) is too strong a condition and it is appropriate to introduce the concept of *diffidence* to characterize this weaker notion of local aversion to zero-mean lotteries.

Definition 7 *The agent is* diffident *at wealth w if (28) holds.*

If u is differentiable at w, then diffidence at that wealth level corresponds to local concavity of u in that point.

Example 4 As a first application of the diffidence theorem, let us characterize central dominance introduced in Section 5.3. Using the basis $\mathcal{B}_2$ for $\mathcal{U}_2$ as before in Section 5.2 we know that any $u \in \mathcal{U}_2$ can be represented as expectation,

$$u(x) = \int_a^b b(x,\theta)dH(\theta) = \int_a^b xdH(\theta) + \int_a^b \min\{x,\theta\}dH(\theta),$$

given a suitable cdf H over the basis functions $b(\cdot,\theta) \in \mathcal{B}_2$, $\theta \in [\underline{\theta},\bar{\theta}] = \Theta$. Thus,

$$u'(x) = 1 + \mathbf{1}_{\{x \leq \tilde{\theta}\}}.$$

Let $f(\theta) = E\tilde{x}\left(1 + \mathbf{1}_{\{w+\tilde{x}\leq\theta\}}\right)$ and $g(\theta) = -E\tilde{y}\left(1 + \mathbf{1}_{\{w+\tilde{y}\leq\theta\}}\right)$, so that we can rewrite (21) in the familiar form

$$Ef(\tilde{\theta}) = 0 \;\Rightarrow\; Eg(\tilde{\theta}) \leq 0.$$

Assuming without loss of generality that $a \leq \underline{\theta} - w < \bar{\theta} - w \leq b$, Proposition 7 then yields: *$\tilde{x}$ centrally dominates $\tilde{y}$ at w, if and only if there exists a constant λ such that*

$$\int_a^{\vartheta} ydG(y) + E\tilde{y} \geq \lambda \left(\int_a^{\vartheta} xdF(x) + E\tilde{x} \right) \tag{29}$$

[21] Naturally we need to replace all strict inequalities in Definition 2 and Proposition 4 by weak inequalities.

for all $\vartheta \in [\underline{\theta} - w, \bar{\theta} - w]$. For pure risks with $E\tilde{x} = E\tilde{y} = 0$ this condition simplifies to

$$\int_a^\vartheta y dG(y) \geq \lambda \int_a^\vartheta x dF(x) \tag{30}$$

for all $\vartheta \in \Theta - w$. □

Other applications of the diffidence theorem are given in the appendix (alternative proof of Jensen's inequality) and Problem 6.

6.2 Monotone Comparative Statics

Let us now consider a parameter set $\mathcal{T} \subset \mathbb{R}$ and let us examine changes of maximizers

$$a(t) = \arg\max_{\hat{a} \in \mathcal{A}} \varphi(\hat{a}; t), \tag{31}$$

where the objective function $\phi : \mathcal{A} \times \mathcal{T} \to \mathbb{R}$ is here assumed to be at twice continuously differentiable for simplicity.[22] In addition, let the finite-dimensional action set $\mathcal{A}$ be nonempty, compact and convex. We say that the optimization problem (31) exhibits monotone comparative statics, if $a(t)$ is monotonically increasing on the parameter set $\mathcal{T}$. The following proposition[23] guarantees smoothness properties of the maximizer $a(t)$.

Proposition 8 (Maximum Theorem; Berge, 1963) *If* $\varphi(\cdot; t)$ *is continuous for any* $t \in \mathcal{T}$*, then: (i)* $a(t)$ *is compact-valued and upper hemicontinuous (uhc) on* $\mathcal{T}$*. (ii)* $\varphi(a(t); t)$ *is continuous on* $\mathcal{T}$*.*

Proof. See Berge (1963, pp. 115–117).

We thereby say that a (set-valued) mapping $a(t)$ is *upper hemicontinuous* at a point $t_0 \in \mathcal{T}$, if for each open set $\mathcal{S} \supset a(t_0)$ there exists a neighborhood $\mathcal{U}(t_0)$ such that: $t \in \mathcal{U}(t_0) \;\Rightarrow\; a(t) \subset \mathcal{S}$. The mapping $a(t)$ is uhc on the open set $\mathcal{T}$ if it is uhc at every point of $\mathcal{T}$. If $a(t)$ is single-valued, then

[22] The smoothness assumptions are not necessary in the least for the main results to hold. Since main insights are based on lattice-theory everything carries over to discrete optimization, which – however – is beyond the scope of this course.

[23] The maximum theorem is due to Berge (1963, p. 116) and it is *very* important to keep it in mind when solving optimization problems. All conclusions (i.e., uhc of maximizer and continuity of $\varphi(a(t); t)$) still hold if in addition to the version stated here we let the action set $\mathcal{A}(t) \neq \emptyset$ depend on the parameter t as well. The set-valued map $\mathcal{A} : \mathcal{T} \to P(\mathcal{A})$ then needs to be continuous in the topological sense (i.e., pre-image of every open set is open).

uhc is nothing else than continuity. At points of indifference, where $a(t)$ is set valued, the maximized function $\varphi(a(t),t)$ is also continuous, whereas the maximizer might "step over" intermittently without ever really "jumping." The following example clarifies the implications of the maximum theorem.

Example 5 Consider a portfolio problem with ex-post payment under *limited liability.* You can imagine a situation in which an investor (of initial wealth w) has to pay a pre-agreed price $p \geq 0$ *after* an uncertain return $\tilde{r}$ has realized. The investor signed this ex-post payment contract in order to take part in this unique investment opportunity. The question is, how does his optimal investment action $a(p)$ change as the ex-post price p increases. To make things concrete, assume that $\tilde{r} \in \{-1, \rho\}$ with equal probability on the outcomes $r = -1$ (lose all invested money) and $r = \rho > 1$ (note that the expected value of the investment opportunity is positive). Let $w \geq (\ln \rho)/(1+\rho)$ for simplicity and let the investor's utility for money be CARA (cf. page 21), such that for any monetary outcome x his utility is given by $u(x) = -\exp(-x)$. Thus, for any price p the investor solves

$$a(p) \in \arg\max_{\hat{a}\in[0,w]} Eu\left([w - p + \hat{a}\tilde{r}]_+\right),$$

provided that it is small enough so that the expected utility of the investment opportunity is at least $u(w)$. Assume that $p \leq w$, then we can rewrite the investor's problem,

$$a(p) \in \arg\max_{\hat{a}\in[0,w]} -\left(e^{-[w-p-\hat{a}]_+} + e^{-(w-p+\hat{a}\rho)}\right) \subseteq \left\{\frac{\ln \rho}{1+\rho}, w\right\}.$$

The investor is indifferent between the *interior* maximizer $a^i = (\ln \rho)/(1+\rho)$ and *"going for broke"* (i.e., investing everything, $\bar{a} = w$), if and only if

$$1 + e^{-(w-p+w\rho)} = e^{-(w-p)}\left(\rho^{-\frac{1}{1+\rho}} + \rho^{-\frac{\rho}{1+\rho}}\right),$$

so that the solution to the investor's maximization problem (provided he accepts the contract) can be written as

$$a(p) = \begin{cases} (\ln \rho)/(1+\rho), & \text{if } p \leq p_0, \\ w, & \text{if } p \geq p_0, \end{cases}$$

where

$$p_0 = w - \ln\left(\rho^{-\frac{1}{1+\rho}} + \rho^{-\frac{\rho}{1+\rho}} - e^{-w\rho}\right).$$

Note that $a(p)$ is piecewise constant and single-valued for all admissible $p \neq p_0$. At $p = p_0$ the investor is indifferent ($a(p_0)$ contains two elements) and "smoothly steps over" from one maximizer to the other (cf. Figure 7). The maximized expected utility $Eu([w-p+a(p)\tilde{r}]_+)$ is therefore continuous in p as asserted by Berge's maximum theorem. $\square$

Monotone Comparative Statics Under Certainty. Necessary and sufficient conditions on φ so that $a(t)$ exhibits monotone comparative statics (MCS) have been provided by Milgrom and Shannon (1994). Before considering their result, let us briefly examine the *implicit function* approach that can be used if the maximand is sufficiently smooth (twice continuously differentiable) in its arguments and the maximizer is single-valued. The first-order necessary optimality condition for (31) is

$$\varphi_a(a(t);t) = 0, \tag{32}$$

for all $t \in \mathcal{T}$, so that by differentiating with respect to t we obtain that

$$a'(t) = -\frac{\varphi_{at}(a(t);t)}{\varphi_{aa}(a(t);t)}, \tag{33}$$

provided that $\varphi_{aa} \neq 0$. If φ is strictly concave in a, then a necessary and sufficient condition for $a(t)$ to exhibit MCS is that

$$\varphi_{at}(a(t);t) \geq 0. \tag{34}$$

The last condition appears in many economic papers, although it does require the rather strong assumption that φ be concave. The beauty of Milgrom and Shannon's result is that it does not depend on such assumptions at all! All that matters for MCS is a weak complementarity property between the arguments of the objective function, typically referred to as single-crossing property or more generally as *quasi-supermodularity.*

Before we report the full characterization of MCS under certainty, let us formulate and prove a simpler (here only one-dimensional result) due to Topkis (1968, p. 55). This result uses the concept of *increasing differences*, which can be readily interpreted in terms of *complementarity.*

Definition 8 *The function* $\varphi : \mathcal{A} \times \mathcal{T} \to \mathbb{R}$ *has* increasing differences *in* (a,t)*, if for any* $(\hat{a},\hat{t}) \geq (a,t)$*:* $\varphi(\hat{a},\hat{t}) - \varphi(a,\hat{t}) \geq \varphi(\hat{a},t) - \varphi(a,t)$*.*

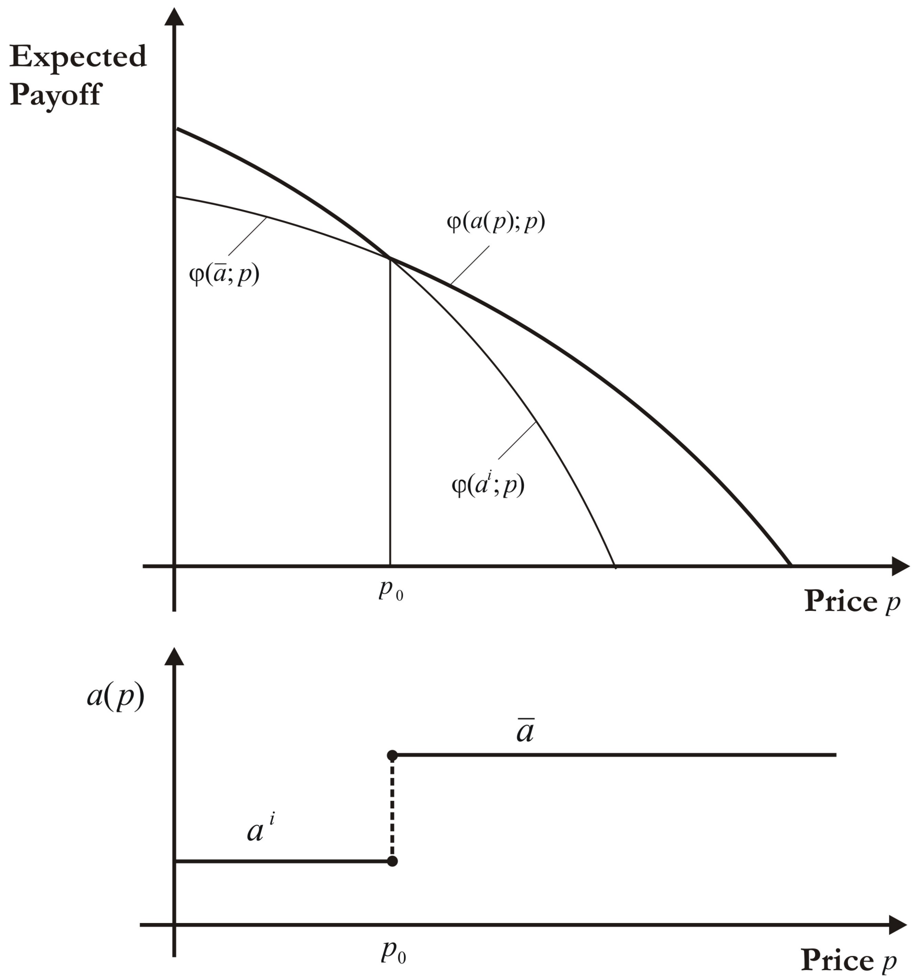

Figure 7: Maximizer $a(p)$ and maximized objective function $\varphi(a(p), p) = Eu\left([w - p + a(p)\tilde{r}]_+\right)$ in Example 5.

The fact that the objective function has increasing differences means that the incremental gain from choosing $\hat{a}$ over a increases in t. Similarly, the incremental gain from having a high parameter $\hat{t}$ instead of t increases in the decision maker's action, since $\varphi(\hat{a},\hat{t}) - \varphi(\hat{a},t) \geq \varphi(a,\hat{t}) - \varphi(a,t)$ by symmetry of Definition 8. Note that the sets $\mathcal{A}$ and $\mathcal{T}$ do merely have to be (partially) ordered for increasing differences to make sense; they could be discrete for instance. It is clear that if φ is twice continuously differentiable that *φ has increasing differences if and only if $\varphi_{at} \geq 0$.* The property of increasing differences is thus directly related to MCS by (34), as will become clear with the following result.

Proposition 9 (Topkis, 1968) *Let $\varphi : \mathcal{A} \times \mathcal{T} \to \mathbb{R}$ have increasing differences and be upper semicontinuous in a. Then $a(t)$ exists for all $t \in \mathcal{T}$ and possesses a smallest element $\underline{a}(t)$ and a largest element $\bar{a}(t)$. Furthermore, $\underline{a}(t)$ and $\bar{a}(t)$ are increasing in t.*

Proof. Given a particular $t \in \mathcal{T}$, select an increasing sequence $(a^k(t))_{k=0}^{\infty} \subset a(t)$ with $a^{k+1} \geq a^k$. Define $\bar{a} = \lim_{k\to\infty} a^k$, which is finite for $\mathcal{A}$ is bounded. Since φ is *upper semicontinuous* at $\bar{a}$, for any $\varepsilon > 0$ there exists a neighborhood $\mathcal{U}(\bar{a})$ such that: $a \in \mathcal{U}(\bar{a}) \;\Rightarrow\; \varphi(a;t) \leq \varphi(\bar{a};t) + \varepsilon$. Hence, there is a $K > 0$ such that for all $k \geq K$: $a^k \in \mathcal{U}(\bar{a})$ and thus $\varphi(a^k;t) \leq \varphi(\bar{a};t) + \varepsilon$. Thus, for any $a \in \mathcal{A}$: $\varphi(a^k,t) \geq \varphi(a,t) \;\Rightarrow\; \varphi(\bar{a},t) \geq \varphi(a,t)$. (If $\varphi(\bar{a},t) < \varphi(a,t)$ then just select $\varepsilon > 0$ small enough, so that necessarily $\varphi(a^k,t) < \varphi(a,t)$ for all $k \geq K(\varepsilon)$.) Hence $\bar{a}(t) \in a(t)$, i.e., $\bar{a}(t)$ is indeed a maximizer. The proof proceeds analogously to show that $\underline{a}(t) \in a(t)$. We will now show the monotonicity of $\bar{a}(t)$. For this, let $\hat{t} \geq t$ and $a \in a(t)$, $\hat{a} \in a(\hat{t})$. Then $\varphi(a;t) - \varphi(\min\{a,\hat{a}\};t) \geq 0$, since $a \in a(t)$. If $a \leq \hat{a}$, then the LHS of the last inequality is in fact zero. If on the other hand $a \geq \hat{a}$, then the inequality becomes $\varphi(a;t) - \varphi(\hat{a};t) \geq 0$. As a result of both of these cases: $\varphi(\max\{a,\hat{a}\};t) - \varphi(\hat{a};t) \geq 0$ and thus by the increasing-differences property of φ:

$$\varphi(\max\{a,\hat{a}\};\hat{t}) - \varphi(\hat{a};\hat{t}) \geq \varphi(\max\{a,\hat{a}\};t) - \varphi(\hat{a};t) \geq 0.$$

We have therefore shown that $\max\{a,\hat{a}\}$ maximizes $\varphi(\cdot;\hat{t})$. If we now set $a = \bar{a}(t)$ and $\hat{a} = \bar{a}(\hat{t})$, then $\bar{a}(\hat{t}) \geq \bar{a}(t)$. One can conclude in a similar manner that $\underline{a}(\hat{t}) \geq \underline{a}(t)$, which concludes the proof. ■

Note that for Proposition 9 to hold, in addition to $\mathcal{A}$ being compact the sets $\mathcal{A}$ and $\mathcal{T}$ merely have to be partially ordered. The following results based on the concept of (quasi-)supermodularity are technically more involved as they use lattice-theoretic methods, which are beyond the scope of this course. For more details, see Topkis (1998, Chapter 2).

Definition 9 *Let $g : \mathcal{T} \to \mathbb{R}$ and $\mathcal{T} \subset \mathbb{R}$. (i) g satisfies* single-crossing in t *(SC1), if there exists $t_0 \in \mathcal{T}$ such that $(t - t_0)(g(t) - g(t_0)) \geq 0$ for all $t \in \mathcal{T}$. (ii) Let $\mathcal{A} \subset \mathbb{R}^n$, where $\mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$ with $\mathcal{A}_k \subset \mathbb{R}$ for all $k = 1, \ldots, n$. The function $\varphi : \mathcal{A} \times \mathcal{T} \to \mathbb{R}$ satisfies* single-crossing in (a, t) *(SC2), if for all $a_H > a_L$: $g(t) = \varphi(a_H; t) - \varphi(a_L; t)$ satisfies SC1. (iii) The function $h : \mathcal{A} \to \mathbb{R}$ is* quasi-supermodular *if it is SC2 in (a_k, a_j) for all $k, j \in \{1, \ldots, n\}$ with $k \neq j$.*

A function is quasi-supermodular if it satisfies a single-crossing property (SC2) in all possible variable pairs. The following result completely characterizes MCS.

Proposition 10 (MCS; Milgrom and Shannon, 1994) *Let $\mathcal{A} \subset \mathbb{R}^n$ be a product set as in Definition 9 (ii). The maximizer $a(t)$ in (31) is increasing if and only if the objective function $\varphi : \mathcal{A} \times \mathcal{T} \to \mathbb{R}$ is quasi-supermodular.*

Proof. See Milgrom and Shannon (1994).

Since it is in most situations not that easy to verify quasi-supermodularity, we are interested in simpler sufficient conditions for MCS. A property that can be checked more easily is supermodularity. Its definition is most natural using the $\vee$ ("join") and $\wedge$ ("meet") operators. For any two vectors $x, y \in \mathbb{R}^n$ and function $h : \mathbb{R}^n \to \mathbb{R}$ we define $x \vee y = \inf\{z \in \mathbb{R}^n : z \geq x \textbf{ and } z \geq y\}$ and $x \wedge y = \sup\{z \in \mathbb{R}^n : z \leq x \textbf{ and } z \leq y\}$.

Definition 10 *Let $h : \mathbb{R}^n \to \mathbb{R}$. (i) The function h is* supermodular,[24] *if for any $x, y \in \mathbb{R}^n$: $h(x \vee y) + h(x \wedge y) \geq h(x) + h(y)$. (ii) If $h > 0$ the function h is* log-supermodular, *if $\log h$ is supermodular,[25] i.e., if for any $x, y \in \mathbb{R}^n$: $h(x \vee y)h(x \wedge y) \geq h(x)h(y)$.*

If h is twice continuously differentiable, then h is supermodular if and only if all cross-partial derivatives (with respect to *different* variables) are nonnegative. For $n = 2$, supermodularity and the increasing-differences property

[24] The function h is *submodular*, if the inequality is reversed. If h is both supermodular and submodular, it is called a *valuation*.

[25] For the definition to make sense, it is enough that $h \geq 0$ and the inequality is satisfied.

in Definition 8 are equivalent.

Monotone Comparative Statics Under Uncertainty. Let us now focus on the situation, when the decision maker's objective function $\varphi : \mathcal{A} \times \mathcal{T} \to \mathbb{R}$ can be formulated as expected utility, in the form (cf. (1) on page 3)

$$\varphi(a;t) = EU(a;t) = \int_\Omega u(a,\omega) f(\omega;t) d\omega, \tag{35}$$

where u is a utility function representation of the decision maker's von Neumann-Morgenstern preferences and $f : \Omega \times \mathcal{T} \to \mathbb{R}_+$ is a (parametrized) pdf.[26] We know by Proposition 10 that the maximizer $a(t)$ in (31) exhibits MCS if and only if φ is quasi-supermodular. Athey (2002) provides necessary and sufficient conditions for the primitives u and f in (35) so that φ becomes quasi-supermodular. We focus on sufficient conditions, as the necessary conditions (based on the construction of appropriate indicator functions) is not easy to operationalize for any given model.

Proposition 11 (MCS Under Uncertainty; Athey, 2002) *Let the decision maker's objective function $\varphi : \mathcal{A} \times \mathcal{T} \to \mathbb{R}$ be represented in the form (35). Then each of the following conditions is sufficient for the maximizer $a(t)$ in (31) to exhibit MCS: (i) $u \geq 0$ is log-supermodular and f is log-supermodular. (ii) u satisfies SC2 and f is log-supermodular.*

Proof. See Athey (2002): in particular Lemmas 4,5 and Theorems 1,2.

7 Notes

There are a number of criticisms related to the expected utility paradigm. Those criticisms, such as the framing effect, the Ellsberg and Allais paradoxes, endowment effect to mention a few are often founded on quite robust empirical evidence. The edited volume by Kahneman and Tversky (2000) provides an excellent introduction to the field of "behavioral economics." There is currently substantial research activity in this area, especially in

[26]Note that the representation (35) does not cover the more general case (3) under decomposable consequences (on page 4). The reason is that the action a in (35) does not enter the pdf directly. Nevertheless, sometimes it is possible to still use Athey's results such as in Example 5 (cf. Problem 8) for instance.

providing rational foundations for empirical observations by formally integrating decision biases into choice models. The second half of these notes owes a lot to Pratt (1964) and Gollier (2001). For a more detailed treatment of stochastic orders, see Shaked and Shantikumar (1994) or more recently Müller and Stoyan (2002). Topkis (1998) is an excellent resource for details on MCS; there is also an unfinished (and unpublished) research monograph by Athey, Milgrom, and Roberts that can currently be downloaded from Susan Athey's Stanford homepage.

8 Problems

Problem 1 (Optimization Practice) (i) Carefully characterize optimal actions for the decision problems outlined in Example 1 and Example 2. Justify any simplifying assumptions you need to make in order to arrive at such representations. (ii) [Sensitivity Analysis] Can you make statements about how optimal actions change if problem parameters shift?

Problem 2 (Utility Representation) Mr. A is not only interested in his own wealth, w_A, but also in Ms. B's wealth, w_B. Mr. A *always* strictly prefers more collective wealth, $w_A + w_B$ to less. In choices that are invariant in collective wealth, he strictly prefers more equitable allocations of wealth between Mr. A and Ms. B. Try to find a utility-representation of Mr. A's preferences. (i) Assume that both Mr. A's and Ms. B's potential wealth lie in the bounded interval $[0, \bar{w}]$ for some positive $\bar{w} < \infty$. Suppose further that there exists a smallest finite money increment $\varepsilon > 0$ between two different wealth levels (e.g., $\varepsilon = 1$ cent). Is it possible to represent Mr. A's utility function? Why or why not? If yes, provide an explicit expression for $u(w_A, w_B)$ — is it unique? (ii) Let $\bar{w} \to \infty$ and/or $\varepsilon \to 0+$. Is it possible to represent Mr. A's utility function? Why or why not? If yes, provide an explicit expression for $u(w_A, w_B)$ — is it unique? (iii) How would you describe the effect of finiteness comparing (i) and (ii)?

Problem 3 (Machina Triangle) Consider a money lottery $\tilde{x} \in \mathcal{X}$, which can have three real-valued outcomes, $\mathcal{X} = \{x_1, x_2, x_3\}$ with $x_1 < x_2 < x_3$. Let p_i be the probability of outcome x_i. Since $p_3 = 1 - p_1 - p_2$, each lottery can be described completely by the tuple $p = (p_1, p_2)$. The set of all admissible tuples $\mathcal{P} = \{(p_1, p_2) \in \mathbb{R}^2_+ : p_1 + p_2 \leq 1\}$ is called the "Machina triangle." The lottery $p = (.2, .4)$ is thus a point in the Machina triangle. (i) Mark the set $\mathcal{L}_1$ of all lotteries that first-order stochastically dominate p. (ii) Mark the set $\mathcal{L}_2$ of all lotteries that second-order stochastically dominate p. (iii) Mark the set of all lotteries $\mathcal{L}_{\text{MPR}}$ that dominate p in the sense of the monotone-probability-ratio stochastic order. (iv) Mark the set $\mathcal{L}_{\text{MLR}}$ of all lotteries that stochastically dominate p in the sense of the monotone-likelihood-ratio stochastic order. (v) Examine and prove all possible inclusion relationships between the different sets $\mathcal{L}_1$, $\mathcal{L}_2$, $\mathcal{L}_{\text{MPR}}$, and $\mathcal{L}_{\text{MLR}}$ (e.g., $\mathcal{L}_1 \subsetneq \mathcal{L}_2$). (vi) Let u be a utility-representation of a decision maker's preferences, which satisfy the von Neumann-Morgenstern axioms. What does the set $\mathcal{L}_0$ of all lotteries that provide the *same expected utility* to the decision maker as p look like? Describe $\mathcal{L}_0$ in general and draw it for a numerical example (i.e., specific values for $u(x_i)$, $i = 1, 2, 3$) that you make up.

Problem 4 (Third-Order Stochastic Dominance) Let

$$\mathcal{U}_3 = \{u \in C^3[a, b] : u'' \leq 0 \leq u', u'''\}$$

for some real constants $a < b$. (i) Show that $\mathcal{U}_3$ is convex and find an appropriate basis $\mathcal{B}_3$. (ii) Establish the corresponding stochastic dominance order, $\preceq_3$. (iii) Mark the set $\mathcal{L}_3$ of all lotteries that third-order stochastically dominate p in Problem 3. [Hint: for a classical treatment of third-order stochastic dominance, see Whitmore's original 1970 paper.]

Problem 5 (Value of a Call Option) A risk-neutral agent owns a European call option on an asset. The value of the underlying asset at the exercise date is described by a random variable $\tilde{x}$ that takes values in $[a, b]$. The strike price of the option is given and equal to y. (i) Show that the agent prefers an *increase in risk* in the value of the underlying asset. (ii) What happens if the agent owns more than one call option on the same asset with different exercise prices?

Problem 6 (Covariance Rule) Let $f : \mathbb{R} \to \mathbb{R}$ be an increasing differentiable function (such that $f' \geq 0$). Let $\tilde{x}$ be an arbitrary random variable. (i) Show that $Ef(\tilde{x})g(\tilde{x}) \leq Ef(\tilde{x})Eg(\tilde{x})$ if and only if the differentiable function $g : \mathbb{R} \to \mathbb{R}$ is decreasing (i.e., $g' \leq 0$). (ii) If $Ef(\tilde{x}) = x_0$ (for otherwise arbitrary $\tilde{x}$), then the inequality in (i) holds if and only if there is a real constant $\lambda(x_0)$ such that g satisfies the following single-crossing condition

$$(g(x) - \lambda(x_0))(x_0 - x) \geq 0.$$

This result is related to the *covariance rule*, $\text{cov}(\tilde{y}, \tilde{z}) = E\tilde{y}\tilde{z} - E\tilde{y}E\tilde{z}$, by setting $\tilde{y} = f(\tilde{x})$ and $\tilde{z} = g(\tilde{y})$. [Hint: use the diffidence theorem.]

Problem 7 (Risk Aversion vs. Diffidence) Consider the utility functions $u(x) = \min\{1, x\}$ and $v(x) = \sqrt{x}$ for $x \geq 0$. (i) Is one more risk averse than the other? (ii) Is one more diffident than the other; if yes, where?

Problem 8 (Monotone Comparative Statics) Examine the comparative statics in Examples 2 and 5. In Example 5 assume that $\Omega = \mathbb{R}$ and that returns are distributed according to a pdf $f : \mathbb{R} \to \mathbb{R}_+$. Discuss this example for a larger class of utility functions u than CARA. [Hint: it may sometimes be useful to change variables.]

References

Anscombe, F.J., Aumann, R.J. (1963) "A Definition of Subjective Probability," *Annals of Mathematical Statistics*, Vol. 34, No. 1, pp. 199–205.

Athey, S. (2002) "Monotone Comparative Statics Under Uncertainty," *Quarterly Journal of Economics*, Vol. 117, No. 1, pp. 187–223.

Aubin, J.-P. (1998) *Optima and Equilibria: An Introduction to Nonlinear Analysis*, Springer, Berlin.

Berge, C. (1963) *Topological Spaces*, Oliver and Boyd, Edinburgh, UK. Reprinted by Dover Publications, Mineola, NY, in 1997. [French original: *Espaces Topologiques, Fonctions Multivoques*, Dunod, Paris, 1959.]

Fishburn, P.C. (1970) *Utility Theory for Decision Making*, Wiley, New York, NY.

Gollier, C. (2001) *The Economics of Risk and Time*, MIT Press, Cambridge, MA.

Gollier, C., Kimball, M.S. (1997) "New Methods in the Classical Economics of Uncertainty: Comparing Risks," Working Paper, University of Michigan.

Halmos, P.R. (1950) *Measure Theory*, Van Nostrand, New York, NY.

Herstein, I.N., Milnor, J. (1953) "An Axiomatic Approach to Measurable Utility," *Econometrica*, Vol. 21, No. 2, pp. 291–297.

HIRSHLEIFER, J., RILEY, J.G. (1992) *The Analytics of Uncertainty and Information*, Cambridge University Press, Cambrigde, UK.

JEWITT, I. (1986) "A Note on Comparative Statics and Stochastic Dominance," *Journal of Mathematical Economics*, Vol. 15, No. 3, pp. 249–254.

KAHNEMAN, D., TVERSKY, A. (EDS.) (2000) *Choices, Values, and Frames*, Cambridge University Press, Cambridge, UK.

KIRILLOV, A.A., GVISHIANI, A.D. (1982) *Theorems and Problems in Functional Analysis*, Springer, New York, NY.

KNIGHT, F. (1921) *Risk, Uncertainty and Profit*, Houghton Mifflin, Boston, MA.

KREPS, D. (1988) *Notes on the Theory of Choice*, Westview Press, Boulder, CO.

MAS-COLELL, A., WHINSTON, M.D., GREEN, J.R. (1995) *Microeconomic Theory*, Oxford University Press, Oxford, UK.

MILGROM, P., SEGAL, I. (2002) "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, Vol. 70, No. 2, pp. 583–601.

MILGROM, P., SHANNON, C. (1994) "Monotone Comparative Statics," *Econometrica*, Vol. 62, No. 1, pp. 157–180.

MORDUKHOVICH, B.S. (1992) "Sensitivity Analysis in Nonsmooth Optimization," in: Field, D.A., Komkov, V., *Theoretical Aspects of Industrial Design*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, pp. 32–46.

MÜLLER, A., STOYAN, D. (2002) *Comparison Methods for Stochastic Models and Risks*, Wiley, New York, NY.

OSBORN, D.W. (2004) *Fractional Stochastic Dominance on Normed Spaces*, Ph.D. Dissertation, Stanford University, Stanford, CA.

Pratt, J. (1964) "Risk Aversion in the Small and in the Large," *Econometrica*, Vol. 32, No. 1/2, pp. 122–136.

Rothschild, M., Stiglitz, J. (1970) "Increasing Risk: I. A Definition," *Journal of Economic Theory*, Vol. 2, No. 3, pp. 225-243.

Savage, L.J. (1954) *The Foundations of Statistics*, Wiley, New York, NY.

Schmeidler, D. (1989) "Subjective Probability and Expected Utility without Additivity," *Econometrica*, Vol. 57, No. 3, pp. 571–587.

Shaked, M., Shantikumar, J.G. (1994) *Stochastic Orders and Their Applications*, Academic Press, Boston, MA.

Strulovici, B.H., Weber, T.A. (2008) "Monotone Comparative Statics: A Geometric Approach," *Journal of Optimization Theory and Applications*, Vol. 137, No. 3, pp. 641–673.

Topkis, D.M. (1968) *Ordered Optimal Solutions*, Ph.D. Dissertation, Stanford University, Stanford, CA.

Topkis, D.M. (1998) *Supermodularity and Complementarity*, Princeton University Press, Princeton, NJ.

Von Neumann, J., Morgenstern, O. (1944) *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ.

Whitmore, G.A. (1970) "Third-Degree Stochastic Dominance," *American Economic Review*, Vol. 60, No. 3, pp. 457–459.

Appendix

Let us now formulate and prove *Jensen's inequality*, which has been instrumental in Section 4. The proof can be given as a straightforward application of the diffidence theorem discussed in Section 6.

Proposition 12 (Jensen's Inequality) *The function $u : \mathbb{R} \to \mathbb{R}$ is concave if and only if for any real random variable $\tilde{x}$ with support in $[a, b]$: $Eu(\tilde{x}) \leq u(E\tilde{x})$.*

Proof. Let us first rewrite Jensen's inequality in the following form:

$$E\tilde{x} = \mu \;\Rightarrow\; Eu(\tilde{x}) \leq u(\mu) \tag{36}$$

for any real random variable $\tilde{x}$ (for which a finite mean μ exists). Using the diffidence theorem (Proposition 7) with $f(x) = x - \mu$ and $g(x) = u(x) - u(\mu)$ implies that (36) is equivalent to

$$u(x) - u(\mu) \leq \lambda(x - \mu)$$

for some constant λ and all x. In other words, $u(x)$ lies (weakly) below the tangent $u(\mu) + \lambda(x - \mu)$ on u in μ. And since this is true *for all* $\mu \in [a, b]$, it is equivalent to u being concave on $[a, b]$. ■