UNIVERSITAT DE BARCELONA

Facultat de Matemàtiques i Informàtica

GRAU DE MATEMÀTIQUES

Treball final de grau

The Von Neumann-Morgenstern Theory and Rational Choice

Autor: Daniel Juan Carreño

Director: Dr. José Manuel Corcuera Valverde

Realitzat a: Departament de Matemàtiques i Informàtica

Barcelona, June 21, 2020

Abstract

Expected Utility Theory (EUT) is an axiomatic theory of choice under risk that has held a central role in economic theory since the 1940s. Throughout this work we will give a thorough description of the von Neumann-Morgenstern theory, which is the cornerstone of EUT and Games Theory. Using the theory and its axiomatisation, we will review its main drawbacks in order to give a proper model of Rational Choice. Furthermore, we will briefly introduce Risk Management, which is a crucial process used to make investment decisions. The process consists in identifying and analysing the amount of risk involved in an investment, and either accepting that risk or mitigating it.

2010 Mathematics Subject Classification. 62P05, 91B08, 91B16, 91B82, 91G70

Acknowledgements

Primer de tot, m'agradaria donar les gràces al Dr. José Manuel Corcuera per donar-me l'oportunitat de treballar en aquest tema, per la seva dedicació al projecte i per la seva ajuda.
Voldria agrair a la meva família el suport durant aquests anys. Gràcies per proporcionar-me tots el recursos necessaris per a poder seguir estudiant i per donar-me força i ànims constants.
Gràcies a tots els meus amics, pel recolzament i comprensió en aquest període de quatre anys.
I per últim, gràcies Júlia per portar llum quan tot era fosc.

Contents

Introduction

Humans are constantly facing decisions that involve risk. In these situations, we try to analyse the probabilities of each possible alternative in order to choose those that yield the greatest profit or have the best outcome according to our preferences. *Expected Utility (EU) theory* addresses the decision-making problem by modelling rational choice. Each possible outcome is assessed in terms of its *utility* and associated to a utility function — its numerical representation—. The von Neumann-Morgenstern utility hypothesis provides the necessary and sufficient conditions under which the Expected Utility theory holds. These conditions are widely accepted to be axioms for the rational choice, however, it has been proved that some of these conditions are violated by real decision-makers. In chapter 3 we will review the most relevant of these violations, and use it to give an extension to the model so it can accommodate the irrational behaviour.

The introduction of the concept of expected utility is usually attributed to Daniel Bernoulli [5]. He arrived at this concept as a resolution of the so-called *St. Petersburg paradox*, for which a *fair* coin is flipped until the first time heads up. If this is at the kth flip, then the gambler receives 2^k dollars. The question that arose from that is how much to pay for participation in this gamble. Bernoulli suggested that the gambler's goal is not to maximise his expected gain, but to maximise the expectation of the logarithm of the gain. The idea that *homo economicus* considers the expected utility of the gamble, and not the expected value, is a cornerstone of Expected Utility theory.

There are few explicit EU calculations in economics before von Neumann-Morgenstern (vNM) [24] who chose to determine the utility value of a randomised strategy in this mathematically convenient way. Like Bernoulli, vNM are concerned with the case in which probabilities are part of the decision problem.

In the first chapter we will introduce preferences and its utility representation, giving also a brief introduction to the choice functions, another approach to modelling individual choice behaviour. Chapter two is the core of this work in which we will thoroughly go through EU theory and decision under risk. As we said before, in chapter three we will review the main violations of the theory, and extend it to the Subjective EU theory, which will be the base for the last chapter. In which robust representation of risk measures are used to model behaviour towards risk in the monetary market. To finish this work, we will give the foundation of the Conic Finance theory, which is based on the theory related to risk measures.

Chapter 1

Preference and utility

In this first chapter we begin by introducing the theory of the individual decision-making. The decision problem starts with a set of possible mutually exclusive alternatives from which the individual must choose. We will denote this set of alternatives by $\mathcal{X}$. One approach to modelling individual choice behaviour is the *preference relation*, which treats the decision maker's tastes and the second approach is the choice function, which treats the individual's choice behaviour. We will establish some rules and axioms to understand how these work on the first two sections of the chapter. To do so, we will mainly follow Mas-Colell [23] and Föllmer and Schied [14]. Besides, in the third section, we will give an introduction to the representation of the preferences using utility functions, which is the baseline to start the study on the expected utility and the collective decision-making.

1.1 Preference relations

Let $\mathcal{X}$ be a non-empty set of possible choices and $x, y \in \mathcal{X}$. When an economic agent faces two possible choices, the decision will be made according to its preferences. This can be formalised as follows:

Definition 1.1. *A preference order (or relation) on $\mathcal{X}$ is a binary relation $\succ$ with the following properties.*

- *Asymmetry: If $x \succ y$, then $y \nsucc x$.*
- *Negative transitivity: Let $z \in \mathcal{X}$, if $x \nsucc y$ and $y \nsucc z$, then $x \nsucc z$.*

Remark 1.1. *If $x \succ y$, we say that x is strictly preferred to y.*

Remark 1.2. *A preference relation is asymmetric if, and only if, it is antisymmetric (if $x \succ y$ and $y \succ x$ then $x = y$) and irreflexive ($x \succ x$ never holds).*

While the asymmetry just states an order between choices, negative transitivity states that if there is a preference between two choices, then if we add a third choice, there is still a choice which is the least preferable or most preferable.

Proposition 1.1. *The binary relation $\succ$ is negatively transitive if, and only if, $x \succ y$ implies that, for all $z \in \mathcal{X}$, $x \succ z$ or $z \succ y$.*

Proof. Observe that "$x \succ y$ implies that $\forall z \in \mathcal{X}$, $x \succ z$ or $z \succ y$" is equivalent to "if there exists a $z \in \mathcal{X}$ such that $x \not\succ z$ and $z \not\succ y$, then $x \not\succ y$", which is equivalent to the definition of negative transitivity. $\square$

Definition 1.2. *The preference relation $\succ$ on $\mathcal{X}$ induces a weak preference $\succeq$, defined by*

$$x \succeq y \Longleftrightarrow y \not\succ x$$

and an indifference relation $\sim$, defined by

$$x \sim y \Longleftrightarrow x \succeq y \text{ and } y \succeq x$$

Remark 1.3. *Since weak preference and indifference are defined from strict preference, then asymmetry and negative transitivity of $\succ$ are equivalent to the following properties of $\succeq$ and $\sim$:*

i) Completeness of $\succeq$: For every $x, y \in \mathcal{X}$, either $x \succeq y$ or $y \succeq x$ or both.

ii) Transitivity of $\succeq$: If $x \succeq y$ and $y \succeq z$, then $x \succeq z$.

iii) Indifference $\sim$ is:

a) Reflexive: for all x, $x \sim x$.

b) Symmetric: $x \sim y$ implies $y \sim x$.

c) Transitive: If $x \sim y$ and $y \sim z$, then $x \sim z$.

iv) If $x \sim w$, $y \sim z$ and $x \succ y$, then $w \succ y$ and $x \succ z$.

Definition 1.3. *If a binary relation $\succeq$ on a set $\mathcal{X}$ is complete, transitive and asymmetric, we say that the set is a completely (or totally) ordered set.*

In much microeconomic theory, individuals' preferences are assumed to be *rational*. This hypothesis is embodied in the assumptions of *completeness* and *transitivity.* That is, the preference relation $\succeq$ is rational if it is complete and transitive.

The completeness axiom is arguably a very strong assumption in preference and utility theory, since it implies that an individual can state his preferences and that they are temporally stable. However, some objections could be made to it. For a variety of reasons, a rational individual could not be able to choose. For instance, if an individual is given the choice between shooting his dog or shooting his cat, he will balk. On the other hand, given the impossibility of facing the agent with a sequence of pairwise choices in which his preferences appear to cycle, transitivity is also a strong assumption; yet there are several possible objections to it. One of these is the so-called problem of *just perceptible differences*, which refers to the indifference transitivity. On Feldman and Serrano [10] we can see an example of the problem:

Example 1.1. *Let x_1 be a cup of coffee with one grain of sugar in it; x_2 a cup with two grains in it; and so on. It's almost certain that you can't taste the difference between any x_k and x_{k+1}, for any $k \in \mathbb{Z}$, and so you must be indifferent between them. Therefore, you must be indifferent between x_0 and $x_{100000000}$, which is probably false.*

We can also find an example for non-transitive preferences in Fishburn [12], it goes as follows:

Example 1.2. *A professor is about to change jobs. His four most important factors are salary, prestige of the university, department reputation, and location. They are roughly equal, and substantially outweigh other factors. He eventually receives four offers and ranks these under each factor on a scale of* 1 *(minimally acceptable) to* 4 *(couldn't be better) in the following way:*

	Salary	*Prestige*	*Reputation*	*Location*
Offer 1	4	3	1	2
Offer 2	3	2	4	1
Offer 3	2	1	3	4
Offer 4	1	4	2	3

He finds that if one offer is better than another one on at least three factors, then he prefers the former. Therefore he finds the following order: Offer 1 $\succ$ Offer 2 $\succ$ Offer 3 $\succ$ Offer 4. But he also finds that Offer 4 $\succ$ Offer 1, which leads him to a non-transitive preference order as it cycles. We can also see that Offer 1 $\succ$ Offer 3 and Offer 2 $\sim$ Offer 4.

1.2 Choice functions

The second approach to modelling individual choice behaviour, as presented at the beginning of this chapter, is the *choice function*. These functions are used to relate the preferences of an individual to its choice behaviour. Preference is linked to hypothetical choice and choice is linked to revealed preference. We will assume that *choice* is induced from *preference*. Let $\mathcal{F}$ be a family of non-empty subsets of $\mathcal{X}$. As per Suzumura [33], an intended interpretation is that each an every subset

$S \in \mathcal{F}$ denotes the set of available states that could possibly be presented to the agent under an appropriate specification of the environmental conditions.

Definition 1.4. *The pair ($\mathcal{X}$, $\mathcal{F}$) is a choice space.*

Definition 1.5. *A choice function on a choice space ($\mathcal{X}$, $\mathcal{F}$) is a function C defined as*

$$\begin{aligned} C : \mathcal{F} &\to \mathcal{F} \\ S &\mapsto C(S) \subseteq S \end{aligned}$$

We assume that the choice function C is non-empty valued, $C(S) \neq \emptyset$, for all S. When $C(S)$ contains a single element, that element is the individual's choice among the alternatives in S. But it may also contain more than one element, and when it does, the elements of $C(S)$ are the alternatives that the individual might choose. If that is the case, the set $C(S)$ would contain all those alternatives that the individual would choose if he was repeatedly told to face the problem of choosing an alternative from S. As explained by Arrow [3], each element of $C(S)$ is to be preferred to all elements of S not in $C(S)$ and indifferent to all elements of $C(S)$.

Definition 1.6. *The choice function C defined on $\mathcal{F}$ satisfies de* ***weak axiom of revealed preference***[1] *if for any S, S′ $\in \mathcal{F}$ with x, y $\in$ S, S′ we have x $\in$ C(S) and y $\in$ C(S′), then x $\in$ C(S′).*

To put that into words, under the conditions that both x and y are in S and $x \in C(S)$, it is revealed that x is weakly preferred to y. Now if $y \in C(S')$ and $x \in S'$, since x is no worse than y, x should also be among the most preferred things in S'. That is, if $C(\{x,y\}) = \{x\}$, then it is not possible that $C(\{x,y,z\}) = \{y\}$.

Definition 1.7. *Given a choice function C(·) defined on a family $\mathcal{F}$, the* ***revealed weakly preference order*** *$\succeq^*$ is defined by*

$$x \succeq^* y \Longleftrightarrow \textit{there exists } S \subseteq \mathcal{F} \textit{ s.t. } x, y \in S \textit{ and } x \in C(S).$$

Definition 1.8. *Given a preference relation $\succeq$ on $\mathcal{X}$ and a non-empty subset $S \in \mathcal{F}$, the set of acceptable alternatives from S according to $\succeq$ is*

$$C^*(S, \succeq) = \{x \in S : \textit{ there is no } y \in S \textit{ such that } y \succ x\}^2$$

[1]The *weak axiom* was first presented by Samuelson [30] as follows: "if any individual selects batch one over batch two; he does not at the same time select two over one"

[2]In some cases, the set $C(S; \succeq)$ may contain no elements at all. For example, suppose that $\mathcal{X}$ = [0, ∞) with $x \in \mathcal{X}$ representing dollars. Suppose $S = \{1, 2, 3, ...\}$. If you always prefer more money to less, or $x \succ y$ whenever $x > y$, then $C(S; \succeq)$ will be empty.

This definition is just saying that the individual is happy to choose anything that isn't bettered by something else available. That is, the most preferred alternatives in S.

Definition 1.9. *Given a rational preference relation $\succeq$ on $\mathcal{X}$, a non-empty subset $S \in \mathcal{F}$ and a choice function $C(\cdot)$, we say that the rational order $\succeq$ rationalises $C(\cdot)$ relative to $\mathcal{F}$ if, for all $S \in \mathcal{F}$*

$$C(S) = C^*(S, \succeq)$$

Proposition 1.2. *Let $\succeq$ be a rational preference relation. Let $\mathcal{F}$ be a family of non-empty subsets of $\mathcal{X}$, and $C(\cdot)$ a choice function defined on $\mathcal{F}$. Then the choice function $C^*(S, \succeq)$ induced by $\succeq$ satisfies the weak axiom for all $S \in \mathcal{F}$.*

Proof. Suppose we have $x, y \in S$ and $x \in C^*(S, \succeq)$. This implies that $x \succeq y$. Let S' be another subset in $\mathcal{F}$ s.t. x, $y \in S'$, and we have $y \in C^*(S', \succeq)$. Then, for all $z \in S'$, $y \succeq z$, but we know that $x \succeq y$. Therefore, $x \succeq z$ and so $x \in C^*(S', \succeq)$. □

We have now seen that the individual's decision-making doesn't need to be based on a process of introspection and well-thought preferences but can be given an entirely behaviour foundation. The choice behaviour is the result of direct observation. But we have also seen that both approaches are compatible and, in fact, putting them together allows us to give a more thorough description of the individuals' behaviour.

1.3 Numerical representation and utility functions

Once we have described how ordered preferences can represent individuals' behaviour towards a choice problem, we need to make them explicit. We need a form of quantification of these preferences. Quantification facilitates us the search of an optimal, or at least near-optimal, decision. The numerical representation of a preference relation, on a set $\mathcal{X}$, is what we call a *utility function*. A more vulgar approach is that the numbers associated to the different elements of the set $\mathcal{X}$ are called utilities, and a utility function tells us the utility associated to each element $x \in \mathcal{X}$.

Definition 1.10. *A numerical representation of a preference order $\succ$ is a function $U : \mathcal{X} \to \mathbb{R}$ such that, for all x,y $\in \mathcal{X}$,*

$$x \succ y \Longleftrightarrow U(x) > U(y)$$

or equivalently,

$$x \succeq y \Longleftrightarrow U(x) \geq U(y)$$

What 1.10 means is that an agent would make the same choice whether he uses his preference order or his utility function. For example, an individual has the following preference: "I prefer taller basketball players", then the set $\mathcal{X}$ can be conceived as all basketball players and the function U would represent the height of the players.
We must note that utility representations are strictly increasing and, hence, not unique. Any other function that assigns numbers to the alternatives on the set and orders them in the same way the utility function does, will also be a representation for it. Before proving this statement, we have to introduce de binary relation $\succ_U$.

Definition 1.11. *Let U be a real valued function, U: $\mathcal{X} \to \mathbb{R}$. The binary relation $\succ_U$ defined as*

$$x \succ_U y \Longleftrightarrow U(x) > U(y)$$

is a preference relation on $\mathcal{X}$.

It is trivial to see that, indeed, *asymmetry* and *negative transitivity* are both satisfied.

Proposition 1.3. *Utility representation is not unique. In fact, if U is a utility function, then for any strictly increasing function $\phi : \mathbb{R} \to \mathbb{R}$, $\psi \equiv \phi \circ U$ is also a utility function for the same preference relation. The utility function is unique up to strictly increasing transformation.*[3]

Proof. We only need to prove uniqueness. Given two strictly increasing utility functions $U(x)$ and $V(x)$ that represent the same preferences, we look for a strictly increasing function $\phi : \mathbb{R} \to \mathbb{R}$ such that $V(x) = \phi \circ U(x) = \phi[U(x)]$, for any $x \in \mathcal{X}$. We identify ϕ as $\phi(t) = V[U^{-1}(t)]$. More rigorously, for each $t > 0$ within the range $U(\mathcal{X})$ of U, there is $x_t \in \mathcal{X}$ s.t. $t = U(x_t)$. Define $\phi(t) = V(x_t)$. We now show that ϕ is well defined. If there is another $y \in \mathcal{X}$ s.t. $U(y) = U(x_t)$, since U and V represent the same preferences, then $V(y) = V(x_t)$, implying $\phi(t) = V(y)$. That is, although x_t may not be unique for each t, the so-defined value $\phi(t)$ is unique. Then, for any x and y, we have

$$\phi[U(x)] = V[U^{-1}(U(x))] = V(x)\ ,\ \phi[U(y)] = V[U^{-1}(U(y))] = V(y) \qquad (1.1)$$

If ϕ is not strictly increasing, then we can find two values t_1 and t_2, $t_1 < t_2$, s.t. $\phi(t_1) \geq \phi(t_2)$. We can then find two arbitrary elements x_1 and x_2 s.t. $U(x_1) = t_1$ and $U(x_2) = t_2$. We have that $x_2 \succ_U x_1$, but by (2.1), we have

$$V(x_1) = \phi[U(x_1)] = \phi(t_1) \geq \phi(t_2) = \phi[U(x_2)] = V(x_2)$$

i.e., $x_1 \succeq_V x_2$, which contradicts the fact that U and V represent the same preferences. Hence. ϕ must be strictly increasing. $\square$

[3]This proposition and its proof can be found on page 42 of Susheng [37]

Now we want to characterise those preference relations for which there exists a numerical representation. Problems on how to numerically represent the preferences arise when we work with non-countable or infinite sets. Under what assumptions do utility representations exist? All finite sets with a preference relation associated have one, but also all countable, infinite sets do. We will give a three-part proof, following Rubenstein [29], of the existence of numerical representation, starting with a lemma regarding the existence of minimal elements.

Lemma 1.1. *In any finite set $S \subseteq \mathcal{X}$, there is a minimal (maximal) element.*

Proof. By induction on the size of S. If S has only one element, that element is, by completeness, minimal. Now, let S be a subset of cardinality $n+1$ and $x \in S$. The set $S - \{x\}$ has a cardinality of n and by the inductive assumption, it has a minimal element y. If $x \succeq y$, then y is minimal in S. If $y \succeq x$, then by transitivity, for all $z \in S - \{x\}$, $z \succeq x$, and thus x is minimal. □

Proposition 1.4. *If $\succeq$ is a preference relation on a* ***finite*** *set $\mathcal{X}$, then $\succeq$ has a utility representation with values being natural numbers.*

Proof. We will construct, by induction, a sequence of sets. Let S_1 be the subset of elements that are minimal in $\mathcal{X}$. By the lemma, S_1 is not empty. Assume we have constructed the sets $S_1, \ldots, S_k$, $k \in \mathbb{N}$. If $\mathcal{X} = S_1 \cup S_2 \cup \ldots \cup S_k$, we are done. If not, define S_{k+1} to be the set of minimal elements in $\mathcal{X} - S_1 - S_2 - \cdots - S_k$. By the lemma, S_{k+1} is not empty. Since $\mathcal{X}$ is finite, we must be done after $|\mathcal{X}|$ steps. Define $U(x) = k$ if $x \in S_k$. Thus, $U(x)$ is the step number at which x is "eliminated". To verify that U represents $\succeq$, let $a \succ b$. Then, $a \notin S_1 \cup S_2 \cup \ldots \cup S_{U(b)}$ and thus $U(a) > U(b)$. If $a \sim b$, then clearly $U(a) = U(b)$. □

The existence of representation is guaranteed if $\mathcal{X}$ is countable.

Proposition 1.5. *If $\mathcal{X}$ is countable, then any preference order on the set has a utility representation with range (-1,1).*

Proof. Let $\{x_n\}$ be an enumeration of all elements in X. We will construct the utility function by induction. Set $U(x_1) = 0$. Assume that you have completed the definition of the values $U(x_1), \ldots, U(x_{n-1})$ so that $x_k \succeq x_l \Longleftrightarrow U(x_k) \geq U(x_l)$. If x_n is indifferent to x_k for some $k < n$, then assign $U(x_n) = U(x_k)$. If not, by transitivity, all numbers in the non-empty set $\{U(x_k) | x_k \succ x_n\} \cup \{-1\}$ are below all numbers in the non-empty set $\{U(x_k) | x_n \succ x_k\} \cup \{1\}$. Choose $U(x_n)$ to be between the two sets. This guarantees that for any $k < n$ we have $x_n \succeq x_k \Longleftrightarrow U(x_n) \geq U(x_k)$. Thus, the function we defined on $\{x_1, \ldots, x_n\}$ represents the preferences of those elements. To complete the proof that U represents $\succeq$. take any two elements x, $y \in \mathcal{X}$. For some k and l we have $x = x_k$ and $y = x_l$. The above applied to $n = max\{k, l\}$ yields $x_k \succeq x_l \Longleftrightarrow U(x_k) \geq U(x_l)$. $\square$

We have established when a preference relation is representable, but when is it that we cannot represent the preferences? An example of it are the lexicographic preferences.

Example 1.3. *The* ***lexicographic order****, denoted by $\succeq_L$, is defined in the set $\mathcal{X} = [0,1] \times [0,1]$ by $(x_2, y_2) \succeq_L (x_1, y_1)$ if either $x_2 > x_1$, or $x_1 = x_2$ and $y_2 \geq y_1$. Suppose that $\succeq_L$ has a utility representation $U : [0,1] \times [0,1] \to \mathbb{R}$. Given any $x \in [0,1]$, since $(x,1) \succ_L (x,0)$, we must have $U(x,1) > U(x,0)$). Let us denote $I(x) = (U(x,0), U(x,1))$, a non-empty and open interval that contains a rational number q_x. For each $x \in [0,1]$, we would have a rational number q_x, and for $x' \neq x$, as $I(x) \cap I(x') = \emptyset$, $q_{x'} \neq q_x$. That is, we would have an injection of the real numbers into the rational numbers, which is impossible. Hence, $\succeq_L$ has no utility representation.*

We will now have a look at the continuity of preferences and its numerical representation. The basic intuition of the continuity of a preference relation is that if $x \succ y$, small deviations from x or y should not change the ordering.

Definition 1.12. *Let $\mathcal{X}$ be a topological space. A preference relation $\succ$ is called continuous if, for all $x \in \mathcal{X}$*

$$\{y \in \mathcal{X} |\ y \succeq x\} \text{ and } \{y \in \mathcal{X} |\ x \succeq y\}$$

are closed subsets of $\mathcal{X}$.

We will be mainly following Debreu [7] to show when a utility representation is continuous.

Definition 1.13. *The upper (lower) topology on an ordered set $\mathcal{X}$ is the weakest topology for which the set $\{x \in \mathcal{X} | x \succeq y\}$ ($\{x \in \mathcal{X} | x \preceq y\}$) is closed for every y in $\mathcal{X}$.*

Definition 1.14. *A collection $\mathcal{B}$ of elements of a topology on a set $\mathcal{X}$ is called a base (or basis) if*

1. *The union of all elements of $\mathcal{B}$ is equal to $\mathcal{X}$.*
2. *For every two elements U, $V \in \mathcal{B}$, the intersection $U \cap V$ is a union of elements of $\mathcal{B}$.*

Definition 1.15. *A topological space $\mathcal{X}$ is separable if there exists an infinite sequence $\{a_n\}_{n \in \mathbb{N}} \subset \mathcal{X}$ such that, given any point $b \in \mathcal{X}$ and any non-empty open subset U of b, we have $a_i \in U$ for some i. That is to say, $\mathcal{X}$ is separable if it has a countable dense subset.*

Theorem 1.1. *Let $\mathcal{X}$ be a topological space which satisfies at least one of the following two properties:*

- *$\mathcal{X}$ has a countable base of open sets*
- *$\mathcal{X}$ is separable and connected*

Then every continuous preference order on $\mathcal{X}$ admits a continuous numerical representation

The proof for the theorem can be found in Debreu [7]. The theorem together with the following proposition, set the necessary background to prove the existence of a continuous numerical representation of a continuous preference relation.

Proposition 1.6. *Let $\mathcal{X}$ be a completely ordered set. If there is an increasing function $V : \mathcal{X} \to \mathbb{R}$, then there is an increasing function $U : \mathcal{X} \to \mathbb{R}$ that is upper (lower) semi-continuous*[4] *in the upper (lower) topology.*[5]

Theorem 1.2. *Let $\mathcal{X}$ be a topological space with a countable base of open sets. and $\succeq$ a continuous preference order defined on $\mathcal{X}$. If the sets $\{x \in \mathcal{X} | x \succeq y\}$ and $\{x \in \mathcal{X} | x \preceq y\}$ are closed for every $y \in \mathcal{X}$, then there is a continuous, increasing function $U : \mathcal{X} \to \mathbb{R}$ representing the preference order.*

Proof. Firstly, we will construct an increasing function $V : \mathcal{X} \to \mathbb{R}$. We will denote each member of the countable base by O_n, for $n = 1, 2, 3, \ldots$. Let $N(x)$ be the set of integers such that, for an element, $x \in \mathcal{X}$

$$N(x) = \{n | O_n \prec \{x\}\}$$

and set

$$V(x) = \sum_{n \in N(x)} \frac{1}{2^n}$$

If $z \succeq y$, then $N(y)$ is a subset of $N(z)$ and $V(z) \geq V(y)$. Thus V is increasing, for if $z \succ y$, y belongs to the open set $\{x \in \mathcal{X} | x \prec z\}$. For some n, we have $y \in O_n \prec \{z\}$. For that n, we don't have $O_n \prec \{y\}$. Hence, $N(y)$ is a subset of $N(z)$ and $V(z) > V(y)$.
We can now apply the proposition 1.6 and obtain an increasing function $U : \mathcal{X} \to \mathbb{R}$ upper semi-continuous in the upper topology.
Analogously, and by substituting $\succeq$ by $\preceq$ we get the lower semi-continuous function in the lower topology. Since the given topology on $\mathcal{X}$ is stronger than or equal to both these topologies, U is both upper and lower semi-continuous in the given topology and, hence, continuous. □

[4]Let $\mathcal{X}$ be a topological space. A function $f : \mathcal{X} \to \mathbb{R}$ is said to be upper (resp. lower) semi-continuous if, and only if, $\{f(x) \geq a\}$ (resp. $\{f(x) \leq a\}$) is closed for every $a \in \mathbb{R}$.

[5]Proof to this proposition can be found on Debreu [7]

In this first chapter we have presented the environment needed to jump into the Theory of the Expected Utility and Von Neumann-Morgenstern's representations. So far, we have seen some relevant results regarding preference representations. These results guarantee the existence of continuous utility functions representing preferences on ordered sets. In the next chapter we will define the sets on which we will work in a different way, as we will treat them as lotteries with a probability associated.

Chapter 2

Von Neumann-Morgenstern's representation and the EU Theory

John von Neumann and Oskar Morgenstern showed in von Neumann and Morgenstern [24] that under some axioms of rational behaviour, an agent that faces a problem with risky outcomes, will behave in the way that he is maximising the expected value of a utility function related to these outcomes. For example, when an individual buys a stock or a bond, he generally does not know precisely the future value of his investment. Nonetheless, decisions have to be made under such conditions of uncertainty, and the question is whether there is a consistent way of thinking about how to make decisions under these conditions. As explained in Hauser and Urban [16], von Neumann-Morgenstern theory explicitly includes risk in its axiomatic foundations. Risk is modelled by transforming the independent variable by a function that reflects the decision-maker's response to uncertain outcomes. The theory bases the selection of the transformation function on the decision-maker's response to a choice between risky situations and a riskless situation. Let's show that with an example.

Example 2.1. *An individual is presented with a* lottery *for which he has to choose between two cars that cost the same and are equally safe. The first car has a guaranteed mileage of* $5\ l/100km$ *and the mileage of the second car is uncertain, as it is equally likely to be* $6\ l/100km$ *or* $4\ l/100km$*. We can see that the expected value for both cars is the same but the choice depends on the risk aversion of the individual. If he prefers the first car, he is* risk averse. *If he prefers the second one, he is* risk prone. *Otherwise, he is risk neutral. In von Neumann-Morgenstern utility theory,* $U(l/km)$ *is scaled to represent this behaviour.*

When the agent lacks perfect foresight, he assigns probabilities to the different possible options and thus, the set $\mathcal{X}$ can be now identified with a subset $\mathcal{M}$ of the set $\mathcal{M}_1(S, \mathcal{A})$ of all probability distributions on a measurable space $(S, \mathcal{A})$, where

$\mathcal{A}$ is a σ-algebra[1]. An element in $\mathcal{M}$ is called a *lottery*, which is a probability measure, that is to say a lottery μ ($\mu : S \to [0,1]$) is a function that assigns a non-negative number $\mu(x)$ to each *prize* x, where $\sum_{x\in\mathcal{X}} \mu(x) = 1$. We will assume that $\mathcal{M}$ is convex[2].

2.1 Von Neumann-Morgenstern utility functions

Definition 2.1. *A von Neumann-Morgenstern (vNM for short) representation is a numerical representation U of a preference relation $\succ$ on $\mathcal{M}$ if it is of the form*

$$U(\mu) = \int u(x)d\mu(x) \quad \textit{for all } \mu \in \mathcal{M},$$

where u is a real function on S.

The vNM utility function is the mathematical expectation, over the realisations of x, of the values of $u(x)$. The function U is linear in the measurable space $\mathcal{M}$. That is,

$$U(\alpha\mu + (1-\alpha)\lambda) = \alpha U(\mu) + (1-\alpha)U(\lambda)$$

for all μ, $\lambda \in \mathcal{M}$ and $\alpha \in [0,1]$. And we say that U is *affine* on $\mathcal{M}$.

2.2 Expected utility theory

Expected utility theory deals with choosing among acts where the decision-maker does not know for sure which consequence will result from a chosen act. When faced with several acts, the decision-maker will choose the one with the highest *expected utility* (Eatwell, Milgate and Newman [8]). It was first introduced by Daniel Bernoulli [5], and was presented as a resolution to the so-called *St. Petersburg paradox*. The paradox goes as follows: consider a game in which a player bets on how many tosses of a coin will be needed before it turns up heads.

[1] Recall that a σ-algebra is defined as follows. A σ-algebra $\mathcal{A}$ of subsets of $\mathcal{S}$ is a collection of subsets satisfying the following conditions:

a) $\emptyset \in \mathcal{A}$

b) if $B \in \mathcal{A}$ then its complement B^c is also in $\mathcal{A}$

c) If $B_1, B_2, \ldots$ is a countable collection of sets in $\mathcal{A}$ then so is their union $\bigcup_{n=1}^{\infty} B_n$

[2] A set $C \subset \mathbb{R}$ is *convex* if the line segment of any two points in C lies in C, i.e. for all $x_1, x_2 \in C$ and $\theta \in [0,1]$

$$\theta x_1 + (1-\theta)x_2 \in C$$

The initial stake is at 2 euros and it is doubled every time a head appears, that is he receives 2^k euros if the coin comes up heads on the kth toss. What would be a fair price to pay for entering the game? To answer this we need to have a look at what would be the average payout. The expected value of the gain is:

$$\frac{1}{2} \cdot 2 + \frac{1}{4} \cdot 4 + \frac{1}{8} \cdot 8 + \ldots = 1 + 1 + 1 + \ldots = \infty$$

euros, so any finite amount of money can be wagered and the player will still come out ahead on average. Considering solely the expected value of the net change in one's monetary wealth, one should therefore play the game at any price if offered the opportunity. The paradox is the discrepancy between what people seem willing to pay to enter the game and the infinite expected value. The classical resolution of the paradox involves the explicit introduction of the utility function, the expected utility hypothesis, and the presumption of diminishing marginal utility of money, which refers to the fact that the additional benefit a person derives from a given increase of money, diminishes with every increase in the wealth that he already has.
Decision-making under risk considers the special case where the formulation of the problem for the DM includes probabilities for the events, so that he only has to derive the utilities of consequences. Within the framework of expected utility theory, for the evaluation of an act, only its probability distribution over consequences has to be taken into account. In the vNM approach, with probabilities known in advance, one may just describe acts as probability distributions over consequences instead of, for example, as functions from the states of the consequences, which is the Savage approach. As we can see in von Neumann and Morgenstern [24], in expected utility theory under risk, a preference relation $\succ$ is characterised by three axioms. The first one was already presented in the first chapter of this thesis.

Axiom 2.1 (**Weak order**). *A preference order $\succeq$ on $\mathcal{M}$ is complete (that is, for every μ, $\lambda \in \mathcal{M}$ either $\mu \succeq \lambda$ or $\mu \preceq \lambda$) and transitive (that is, for every μ, λ, $\nu \in \mathcal{M}$, $\mu \succeq \lambda$ and $\lambda \succeq \nu$ imply $\mu \succeq \nu$).*

The next axiom is called the *Archimedean* axiom and it imposes a sort of continuity on the preference relation. It requires that no alternative in $\mathcal{M}$ is infinitely more, or less, desirable than any other alternative.

Axiom 2.2 (**Archimedean axiom**). *A preference relation $\succ$ on $\mathcal{M}$ satisfies the Archimedean axiom if for all μ, λ, $\nu \in \mathcal{M}$, $\mu \succ \lambda \succ \nu$, then there exist α, $\beta \in (0,1)$ such that*

$$\alpha\mu + (1-\alpha)\nu \succ \lambda$$

and

$$\lambda \succ \beta\mu + (1-\beta)\nu$$

Since μ is strictly better than λ, then no matter how bad ν is, we can always find a mixture of μ and ν so it is better than λ. And similarly, the other way around. To help understand this, let's consider an example in which we might think that the axiom does not hold. Suppose that μ gives you 100€ for sure, λ 10€ for sure and ν consists of your death. Then you might think that ν is so much worse to λ that no probability α can make $\alpha\mu + (1-\alpha)\nu$ better than λ, even if it is really close to 1. Nonetheless, this is not true. Imagine that you are told you can have 10€ now or, if you choose to drive to a nearby location, then you will get 100€. Every rational being would take the car and go for the hundred euros even though it increases the chances of dying.
The axiom is sometimes called *continuity axiom*, because it can act as a substitute for the continuity of $\succ$ in a suitable topology on $\mathcal{M}$. More precisely, suppose that $\mathcal{M}$ is endowed with a topology for which convex combinations are continuous curves. Then, continuity of the preference order $\succ$ in this topology automatically implies the Archimedean axiom. An alternative approach for this axiom was given by Herstein and Milnor [17], the *mixture continuity* condition. In order to give the description of this alternative condition, we first need to define the following preference sets:

$$A := \{\alpha \in [0,1] \mid \alpha\mu + (1-\alpha)\nu \succeq \lambda\}$$

$$B := \{\alpha \in [0,1] \mid \lambda \succeq \alpha\mu + (1-\alpha)\nu\}$$

for λ, μ, $\nu \in \mathcal{S}$, where $\mathcal{S}$ is a mixture set[3]. Mixture continuity requires that, whenever $\mu, \lambda, \nu \in \mathcal{S}$ with $\mu \succ \lambda$ and $\lambda \succ \nu$, both A and B must be closed.
The third axiom is the independence axiom, it imposes a form of separability on the preference relation.

Axiom 2.3 (Independence). *A preference relation $\succ$ on $\mathcal{M}$ satisfies the independence axiom if, for all μ, λ, $\nu \in \mathcal{M}$ and $\alpha \in (0,1]$,*

$$\mu \succ \lambda \Longleftrightarrow \alpha\mu + (1-\alpha)\nu \succ \alpha\lambda + (1-\alpha)\nu$$

The independence axiom requires that the preference between the lotteries μ and λ be the same whether they are compared directly or embedded in larger, compound, lotteries that are otherwise identical.

[3] A set $\mathcal{S}$ is said to be a mixture set if for any μ, $\lambda \in \mathcal{S}$ and for any α we can associate another element, which we write as $\alpha\mu + (1-\alpha)\lambda$, which is again in $\mathcal{S}$, and where

1. $1\mu + (1-1)\lambda = \mu$
2. $\alpha\mu + (1-\alpha)\lambda = (1-\alpha)\lambda + \alpha\mu$
3. $\beta[\alpha\mu + (1-\alpha)\lambda] + (1-\beta)\lambda = (\alpha\beta)\mu + (1-\beta\alpha)\lambda$

for all μ, $\lambda \in \mathcal{S}$ and all α, β.

Theorem 2.1. *If a preference relation $\succ$ on $\mathcal{M}$ satisfies the three axioms we presented above, then there exists an affine numerical representation U of $\succ$. Furthermore, the function U is unique up to positive affine transformations, i.e., any other affine numerical representation V with these properties is of the form $V = cU + d$ for some $c > 0$ and $d \in \mathbb{R}$.*

Proof. It is easy to see that, if U is an affine function that represents $\succ$, so will be any increasing affine transformation V. Assuming that the axioms hold, we first need a few lemmas to help us prove the theorem. The proof for this lemmas can be found on Herstein and Milnor [17].

Lemma 2.1. *For every μ, $\lambda \in \mathcal{M}$, if $\mu \succ \lambda$, then*

i) for every $\alpha \in (0,1)$, $\mu \succ \alpha\mu + (1-\alpha)\lambda \succ \lambda$

ii) for every $\alpha, \beta \in (0,1)$ s.t. $\alpha > \beta$, $\mu \succ \alpha\mu + (1-\alpha)\lambda \succ \beta\mu + (1-\beta)\lambda \succ \lambda$

Lemma 2.2. *For every μ, $\lambda \in \mathcal{M}$, if $\mu \sim \lambda$, then for every $\alpha \in [0,1]$,*

$$\mu \sim \alpha\mu + (1-\alpha)\lambda \sim \lambda$$

Lemma 2.3. *For every μ, λ, $\nu \in \mathcal{M}$, and every $\alpha \in (0,1)$,*

$$\mu \succeq \lambda \Longleftrightarrow \alpha\mu + (1-\alpha)\nu \succeq \alpha\lambda + (1-\alpha)\nu$$

Lemma 2.4. *Assume that $\mu, \lambda, \nu \in \mathcal{M}$ are such that $\mu \succ \lambda$ and $\mu \succeq \nu \succeq \lambda$. Then there exists a unique $\alpha = \alpha(\mu, \lambda, \nu) \in [0,1]$ such that*

$$\nu \sim \alpha\mu + (1-\alpha)\lambda$$

After having established the background with these lemmas, we can now proceed to give the proof for the theorem. We will construct the affine numerical representation U. We first fix two lotteries λ and ρ, $\lambda \succ \rho$, and define

$$\mathcal{M}(\lambda, \rho) := \{\mu \in \mathcal{M} \mid \lambda \succeq \mu \succeq \rho\}$$

If $\mu \in \mathcal{M}(\lambda, \rho)$, Lemma 2.4 yields a unique $\alpha \in [0,1]$ s.t. $\mu \sim \alpha\lambda + (1-\alpha)\rho$. We set now $U(\mu) := \alpha$. We want to prove that such U is a numerical representation of the preference order $\succ$ on $\mathcal{M}(\lambda, \rho)$, and to do so, we must show that for $\mu, \nu \in \mathcal{M}(\lambda, \rho)$, we have $U(\mu) > U(\nu) \Longleftrightarrow \mu \succ \nu$. Observe that if $U(\mu) > U(\nu)$, by Lemma 2.1, we have

$$\mu \sim U(\mu)\lambda + (1 - U(\mu))\rho \succ U(\nu)\lambda + (1 - U(\nu))\rho \sim \nu$$

and, hence, $\mu \succ \nu$. Contrariwise, if $\mu \succ \nu$, it is clear that we cannot have $U(\nu) > U(\mu)$, so it suffices to rule out the case $U(\mu) = U(\nu)$. Suppose $U(\mu) = U(\nu)$, then by definition we have $\mu \sim \nu$, which contradicts $\mu \succ \nu$. Therefore, U is indeed a numerical representation of $\succ$ restricted to $\mathcal{M}(\lambda, \rho)$. We will show now that $\mathcal{M}(\lambda, \rho)$ is a convex set. Take μ, $\nu \in \mathcal{M}(\lambda, \rho)$ and $\alpha \in [0, 1]$. Then by the independence axiom and Lemma 2.3,

$$\lambda \succeq \alpha\lambda + (1-\alpha)\nu \succeq \alpha\mu + (1-\alpha)\nu$$

Using the same argument, it follows that $\alpha\mu + (1-\alpha)\nu \succeq \rho$, which implies convexity of the set $\mathcal{M}(\lambda, \rho)$. Therefore $U(\alpha\mu + (1-\alpha)\nu)$ is well defined. We have to show that it is also equal to $\alpha U(\mu) + (1-\alpha)U(\nu)$):

$$\begin{aligned}\alpha\mu + (1-\alpha)\nu &\sim \alpha(U(\mu)\lambda + (1-U(\mu))\rho) + (1-\alpha)(U(\nu)\lambda + (1-U(\nu))\rho) \\ &= [\alpha U(\mu) + (1-\alpha)U(\nu)]\lambda + [1 - \alpha U(\mu) - (1-\alpha)U(\nu)]\rho\end{aligned}$$

Then, by the definition of U and the uniqueness of α in Lemma 2.4, we have that

$$U(\alpha\mu + (1-\alpha)\nu) = \alpha U(\mu) + (1-\alpha)U(\nu)$$

Thus, U is indeed well defined and an affine numerical representation of $\succ$.
To end the proof we just have to prove the uniqueness of U up to positive affine transformations. Let V be another affine numerical representation of $\succ$ on $\mathcal{M}(\lambda, \rho)$, and define

$$W(\mu) := \frac{V(\mu) - V(\rho)}{V(\lambda) - V(\rho)}$$

Then W is a positive affine transformation of V, and $W(\rho) = 0 = U(\rho)$ as well as $W(\lambda) = 1 = U(\lambda)$. Hence, affinity of W and the definition of U imply

$$W(\mu) = W(U(\mu)\lambda + (1-U(\mu))\rho) = U(\mu)W(\lambda) + (1-U(\mu))W(\rho) = U(\mu)$$

for all $\mu \in \mathcal{M}(\lambda, \rho)$. Therefore, $W = U$. Finally, we have to show that U can be extended as a numerical representation to the full space $\mathcal{M}$. To this end, we first take $\widetilde{\lambda}$, $\widetilde{\rho} \in \mathcal{M}$ such that $\mathcal{M}(\widetilde{\lambda}, \widetilde{\rho}) \supset \mathcal{M}(\lambda, \rho)$. As we have seen throughout the present proof, there exists an affine numerical representation V of $\succ \in \mathcal{M}(\widetilde{\lambda}, \widetilde{\rho})$, and we may assume that $V(\lambda) = 1$ and $V(\rho) = 0$; otherwise we apply a positive affine transformation to V. By the previous step of the proof, V coincides with U on $\mathcal{M}(\lambda, \rho)$, and so V is a unique consistent extension of U. Since each lottery belongs to some set $\mathcal{M}(\widetilde{\lambda}, \widetilde{\rho})$, the affine numerical representation U can be uniquely extended to all of $\mathcal{M}$. $\square$

There is an important case where such affine numerical representation will already be of vNM form. We formalise this by the next corollary, but first we need to introduce the notion of a *simple probability distribution*. A simple probability distribution is a probability measure μ on S which can be written as a finite convex combination of Dirac measures[4]: there exist $x_1, \ldots, x_N \in S$ and $\alpha_1, \ldots, \alpha_N \in (0,1]$ s.t.

$$\mu = \sum_{i=1}^{N} \alpha_i \delta_{x_i}$$

Corollary 2.1. *Suppose that $\mathcal{M}$ is the set of all simple probability distributions on S and that $\succ$ is a preference relation on $\mathcal{M}$ that satisfies the three axioms. Then there exists a vNM representation U. Moreover, both U and u are unique up to positive affine transformations.*

Proof. Let U be an affine numerical representation. We define $u(x) := U(\delta_x)$, for $x \in S$. If $\mu \in \mathcal{M}$ is of the form $\mu = \alpha_1 \delta_{x_1} + \ldots + \alpha_N \delta_{x_N}$, then affinity of U implies

$$U(\mu) = U(\sum_{i=1}^{N} \alpha_i(\delta_{x_i})) = \sum_{i=1}^{N} \alpha_i U(\delta_{x_i}) = \int u(x) d\mu(x)$$

$\square$

On a finite set, all probability measures are simple. Thus, on a finite set S, any affine numerical representation is already a vNM representation. Problems arise when we work with non-finite sets, in which a vNM representation may not exist. Let's show that with an example.

Example 2.2. *Let $\mathcal{M}$ be the set of all Borel probability measures[5] on $S = [0,1]$, and denote by λ the Lebesgue measure on S. According to the Lebesgue decomposition theorem, every $\mu \in \mathcal{M}$ can be decomposed as*

$$\mu = \mu_S + \mu_a$$

where μ_S is singular with respect to λ and μ_a is absolutely continuous. We define a function $U : \mathcal{M} \to [0,1]$ by

$$U(\mu) := \int x d\mu_a(x)$$

[4] A Dirac measure is a measure δ_x on a set X (with any σ-algebra of subsets of X) defined for a given $x \in X$ and any measurable set $A \subseteq X$ by

$$\delta_x(A) = \mathbb{1}_A(x) = \begin{cases} 0, & x \notin A \\ 1, & x \in A \end{cases}$$

[5] A Borel probability measure on a metric space (X, d) is a map $\mu : \mathcal{B}(X) \to [0, \infty)$ such that

i) $\mu(\emptyset) = 0$

ii) $A_1, A_2, \ldots \in \mathcal{B}$ mutually disjoint $\Rightarrow \mu(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} \mu(B_i)$

iii) $\mu(X) = 1$

Where $\mathcal{B}(X)$ is the smallest σ-algebra containing the open sets.

U is an affine function on $\mathcal{M}$. Hence, U induces a preference order $\succ$ on $\mathcal{M}$ which satisfies the three axioms. But $\succ$ cannot have a vNM representation. Since $U(\delta_x) = 0$ for all x (because λ is singular with respect to all δ_x), the only possible choice for the function u would be $u \equiv 0$. So the preference relation would be trivial in the sense that $\mu \sim \lambda$ for all $\mu \in \mathcal{M}$, in contradiction, for instance, to $U(\lambda) = \frac{1}{2}$ and $U(\delta_{\frac{1}{2}}) = 0$.

In addition to the axioms we have described in this section, we can add another one, the so-called *sure-thing principle*. This one was defined by Savage in 1954 in the famous *Foundations of Statistics* [31]. The axiom goes as follows: "If the person would not prefer f to g, either knowing that the event B prevailed, or knowing that the event $\sim B$ prevailed (denoting the complement of B), then he does not prefer f to g. Moreover (provided that he does not regard B as virtually impossible) if he would definitely prefer g to f, knowing that B prevailed, and, if he would not prefer f to g, knowing that B did not prevail, then he definitely prefers g to f". In our notation, consider first the set $\mathcal{M}_1(S, \mathfrak{B})$ of all probability measures on a separable metric space S, endowed with the σ-field $\mathfrak{B}$ of Borel sets. For $\mu, \nu \in \mathcal{M}$ and $A \in \mathfrak{B}$ such that $\mu(A) = 1$:

$$\delta_x \succ \nu \text{ for all } x \in A \Longrightarrow \mu \succ \nu$$

and

$$\nu \succ \delta_x \text{ for all } x \in A \Longrightarrow \nu \succ \mu$$

This is automatically implied by the existence of a vNM representation.

2.3 Decision under risk

In the previous section, we presented the theory of the expected utility within the framework of the three axioms. The central behavioural concept in expected utility is that of risk aversion, which we introduced in the first example of chapter two. Formally, a situation is said to involve risk if the randomness facing an economic agent can be expressed in terms of specific numerical probabilities. The expected utility concept of risk aversion is a property of attitudes to wealth (the utility function over wealth is concave) rather than of attitudes to risk *per se* (independent of the marginal utility of wealth) (Rabin [27]). In this section, we concentrate on risky alternatives whose outcomes are amounts of money, which represent individual financial assets with their pay-off distribution, at a fixed time, known. The analytical power of the expected utility formulation hinges on specifying the utility function u in such manner that it captures interesting economic attributes to choice behaviour. An individual would always prefer shifting probability mass from lower to higher outcome levels if, and only if, u was an increasing function of x. Hence, we will assume that u is strictly increasing and, also, continuous. Such a shift of probability mass is known as a *first order stochastically dominating shift*.

Definition 2.2. *For any lotteries μ and ν, we say that μ first order stochastically dominates (FOSD) ν if, and only if, the decision-maker weakly prefers μ to ν under every increasing function u:*[6]

$$\int u(x)d\mu(x) \geq \int u(x)d\nu(x).$$

Besides, we will assume that $\mathcal{M}$ contains all point masses δ_x for $x \in S$ and that each $\mu \in \mathcal{M}$ has a well-defined expectation

$$m(\mu) := \int x d\mu(x) \in \mathbb{R}$$

.

Definition 2.3. *The expected value $m(\mu)$ is called the* ***fair price*** *of an asset when μ is the distribution of its pay-off. When μ is the distribution of the payments to be received by an insured party in an insurance contract, $m(\mu)$ is called the* ***fair premium****.*

Let's formally define what risk aversion is.

Definition 2.4. *An individual is* ***risk averse*** *if a certain outcome $\delta_{m(\mu)}$ is always preferred to a risky prospect μ for which $E(\mu) = m(\mu)$. That is $\delta_{m(\mu)} \succ \mu$ unless $\delta_{m(\mu)} = \mu$. If the individual is indifferent between these two lotteries, we say that he is* ***risk neutral****, but we say that he is* ***strictly risk averse*** *if indifference holds only when the two lotteries are the same.*

We can see in Mas-Colell [23] that it follows directly from the definition of risk aversion that the decision-maker is risk averse if, and only if,

$$U(\delta_{m(\mu)}) = u(\int x d\mu(x)) \geq \int u(x)d\mu(x) = U(\mu)$$

This inequality is called *Jensen's inequality* and it is the defining property of a **concave function**. Therefore, in the context of expected utility theory, we see

[6](Mas-Colell [23]) This definition simply states that every individual with increasing utility function prefers μ to ν regardless of his risk preferences. If we take into consideration the distributions of monetary pay-offs, we say that μ *FOSD* ν if, and only if, $\mu(x) \leq \nu(x)$ for every x. This may not be obvious to see. Given these lotteries, denote $\lambda(x) = \mu(x) - \nu(x)$ and suppose $\lambda(\tilde{x}) > 0$ for some $\tilde{x}$. Now define $u(\cdot)$ as $u(x) = 1$ if $x > \tilde{x}$ and $u(x) = 0$ otherwise. We can see that $\int u(x)d\lambda(x) = -\lambda(\tilde{x}) < 0$, so the implied condition is proved. Now, given μ and ν, and considering λ as defined before, we integrate it by parts: $\int u(x)d\lambda(x) = [u(x)\lambda(x)]_0^\infty - \int u'(x)\lambda(x)dx$. Since $\lambda(0) = 0$ and $\lambda(x) = 0$ for large x, the first term is zero. Then. $\int u(x)d\lambda(x) \geq 0 \Longleftrightarrow \int u'(x)\lambda(x)dx \leq 0$. Thus, if $\lambda(x) \leq 0$ for all x and $u(\cdot)$ is increasing, then the second term is lesser or equal than 0, and we have finished.

that *risk aversion is equivalent to the concavity of u*, and that strict risk aversion is equivalent to the strict concavity of u.
Strict concavity means that the marginal utility of money is decreasing, that is to say the utility gain from an extra euro, at any level of wealth x, is smaller than the utility loss of having an euro less. The risk of gaining or losing an euro with even probability is not worth taking the risk.

Definition 2.5. *A preference relation $\succ$ on $\mathcal{M}$ is called monotone if*

$$x > y \Longrightarrow \delta_x \succ \delta_y$$

It is pretty straightforward to see that the fact that function u is increasing implies monotonicity and the other way around. Take $x > y$, then $u(x) = U(\delta_x) > U(\delta_y) = u(y)$ which shows that U is increasing.

We have characterised the utility function used in uncertainty scenarios with risky outcomes.

Definition 2.6. *A function $u : S \to \mathbb{R}$ is called a utility function if it is strictly concave, strictly increasing, and continuous on S.*

Let's now introduce the concept of *certainty equivalent*, a key concept for the analysis of risk aversion.

Definition 2.7. *The **certainty equivalent** of $\mu \in \mathcal{M}$, denoted by $c(\mu)$, is the amount of money for which the individual is indifferent between the lottery μ and the certain amount $c(\mu)$, that is*

$$u(c(\mu)) = \int u(x) d\mu(x)$$

There is indifference between the lottery μ and the sure amount of money $c(\mu)$, that is $\delta_{c(\mu)} \sim \mu$. Since our function u is strictly increasing, every μ has at most one certainty equivalent. Note that risk aversion, in this setting, can be characterised by $m(\mu) \geq c(\mu)$, and

$$m(\mu) > c(\mu) \Longleftrightarrow \delta_{m(\mu)} \neq \mu$$

Since $m(\mu) \geq c(\mu)$ will always hold, we can now define the *risk premium* of μ. The risk premium can be viewed as the amount that an individual would be ready to pay for replacing the asset by its expected value m(μ).

Definition 2.8. *The **risk premium** of $\mu \in \mathcal{M}$ is defined as*

$$\rho(\mu) := m(\mu) - c(\mu)$$

Definition 2.9. *For any fixed amount of money x and a positive number ϵ, the* ***probability premium****, denoted by $\pi(x, \epsilon)$, is the excess in winning probability over fair odds that makes the individual indifferent between the certain outcome x and a gamble between the two outcomes $x + \epsilon$ and $x - \epsilon$. That is*

$$u(x) = (\frac{1}{2} + \pi(x, \epsilon))u(x + \epsilon) + (\frac{1}{2} - \pi(x, \epsilon))u(x - \epsilon)$$

The satisfaction of the inequality $\pi(x, \epsilon) \geq 0$ for all x and $\epsilon > 0$ is also equivalent to risk aversion. We can compute π, for any given x and ϵ, starting from the equation in def. 2.9, as follows:

$$\begin{aligned} u(x) = &\ \frac{1}{2}[u(x+\epsilon) + u(x-\epsilon)] + \pi[u(x+\epsilon) + u(x-\epsilon) \\ \Longrightarrow & u(x) - \frac{1}{2}[u(x+\epsilon) + u(x-\epsilon)] = \pi[u(x+\epsilon) - u(x-\epsilon)] \\ \Longrightarrow & \pi = \frac{u(x) - \frac{1}{2}[u(x+\epsilon) + u(x-\epsilon)]}{u(x+\epsilon) - u(x-\epsilon)} \end{aligned}$$

We can now illustrate the use of the risk aversion concept with a couple of examples of cases on demand of assets and insurance.

Example 2.3 (Demand for a risky asset (Ross [28]))**.** *Suppose that there are two assets; one is a safe asset with a return of 1 euro per euro invested and the other one is a risky asset with a random return of z euros per euro invested. The random return z has a distribution function $\mu(z)$ that we assume satisfies $\int z d\mu(z) > 1$. The individual has an initial wealth of w and it can be divided in any way between the two assets. Consider then, $\lambda \in [0, 1]$ such that $(1 - \lambda)w$ is the amount of money invested in the risky asset and λw in the safe asset. Thus, for any realisation z of the random return, the individual's portfolio pays $(1 - \lambda)wz + \lambda w$. The question is how to choose λ. The answer will depend on μ, w, and the utility function u. We turned the problem into a utility maximisation problem, which is*

$$\max_{\lambda \in [0,1]} \int u((1 - \lambda)wz + \lambda w) d\mu(z) \tag{2.1}$$

Take now $\alpha = (1 - \lambda)w$ so eq. 2.1 is equal to $\int u(w + \alpha(z - 1))d\mu(z)$ and $w \geq \alpha \geq 0$. If the agent is risk-neutral, so that $u(x)$=βx for some constant β, the marginal returns to investment become $\beta w + \beta\alpha(\int z d\mu(z) - 1)$. And these are always positive in our case. The risk neutral investor cares only about the expected rate of return, so he optimally puts all his wealth into the asset with the highest expected return. The objective function is concave in α because the concavity of u implies that
$\int u''(w + \alpha(z - 1))(z - 1)^2 d\mu(z) \leq 0$. If α^* is optimal, it must satisfy the first order condition

$$\psi(\alpha^*) = \int u'(w + \alpha^*(z - 1))(z - 1)d\mu(z) = 0$$

Note that since $\int z d\mu(z) > 1$, *at* $\alpha^* = 0$, *the marginal return to investing a bit more in the risky asset is* $\psi(0) = \int (z-1)u'(w) > 0$. *Hence,* $\alpha^* = 0$ *cannot satisfy the first order condition. Then, we can conclude that the optimal portfolio has* $\alpha^* > 0$, *which implies that* $w(1-\lambda^*) > 0$ *and* $\lambda^* < 1$. *It is illustrated here that if a risk is actuarially favourable*[7]*, then a risk averter will always accept at least a small amount of it.*

Example 2.4 (Demand for insurance). *Consider a strictly risk-averse decision maker who has an initial wealth of* w, *but has a risk of a loss of* Y *euros, with* $w \geq Y \geq 0$. *One unit of insurance costs* q *euros and pays 1 euro if the loss occurs. If* n *units of insurance are bought, the wealth of the individual will be* $w - nq$ *(*nq *is the premium of the insurance) if there is no loss, and* $w - nq - Y + n$ *if the loss occurs. Defining as* λ *the probability of the loss, the DM's expected wealth is*

$$(1-\lambda)(w-nq) + \lambda(w-nq-Y+n) = w - \lambda Y + n(\lambda - q)$$

He has to choose the optimal level of n. *That is*

$$\max_{n \geq 0} \; (1-\lambda)u(w-nq) + \lambda u(w-Y+n(1-q))$$

Denoting by V(n) the objective function and n^* *the optimum amount, the first order condition is*

$$\frac{dV}{dn} = -u'(w-n^*q)(1-\lambda)q + u'(w-Y+n^*(1-q))(1-q)\lambda = 0$$

The first order condition says that the marginal benefit of an extra euro of insurance in the bad state multiplied by the probability of loss, is equal to the marginal cost of the extra euro of insurance in the good state. If we suppose now that q *is actuarially fair*[8]*, meaning* $q = \lambda$, *then the first order condition is*

$$-u'(w-n^*q) + u'(w-Y+n^*(1-q)) = 0$$

Which gives us

$$w - Y + n^*(1-q) = w - n^*q \Longleftrightarrow n^* = Y$$

We can conclude that the DM would take a complete insurance if the insurance was actuarially fair. The DM's final wealth is then $w - \lambda Y$, *regardless of the occurrence of the loss.*

[7] The cost of an asset is less than the asset's expected value.

[8] The price of the insurance policy exactly equals the expected monetary losses.

2.4 Absolute and relative risk aversion

In order to determine the optimal trade-off between the expected gain and the degree of risk, it is useful to quantify the effect of risk on welfare. This is particularly useful when the agent subrogates the risky decision to others, as is the case when we consider public safety policy or portfolio management by pension funds, for example. It is important to quantify the degree of risk aversion in order to help people to know themselves better, and to help them make better decision in the face of uncertainty.

Kenneth J. Arrow and John W. Pratt [3] [26] not only identify the risk aversion with the concavity of a utility function $u(x)$, they also provide a way to measure the degree of concavity and, hence, the strength or intensity of risk aversion. Arrow and Pratt give two related measures and both are extensively used in current economic analysis. Considering a level of wealth x, $u'(x)$ is the marginal utility of wealth, and $u''(x)$ is the rate of change of marginal utility with respect to wealth. Since $u(x)$ is a strictly increasing function, we have $u'(x) > 0$. Suppose that an individual with a level of wealth x_0 is offered a chance to win or lose an amount q at fair odds. He can choose between the sure amount x_0 and an uncertain amount, being it equally probable to get $x_0 + q$ and $x_0 - q$. A risk averter will always prefer the sure amount, translated into expected utility formulation,

$$u(x_0) > \frac{1}{2}u(x_0 + q) + \frac{1}{2}u(x_0 - q)$$

and

$$u(x_0) - u(x_0 - q) > u(x_0 + q) - u(x_0)$$

As we can see, the utility differences corresponding to equal changes in wealth are decreasing as the wealth increases and, thus, $u'(x)$ *is strictly decreasing as x increases* $(\star)$.

One could think that —as the condition $(\star)$ is necessary and sufficient for risk aversion— using its rate of change, $u''(x)$, as a risk measure would be a good idea. But it suffers from a severe formal defect. As we have seen earlier on this work (Theorem 2.1), the utility functions are defined only up to positive affine transformations. Adding a constant to the utility does not change the marginality nor its rate of change, but multiplying $u(x)$ by a constant also multiplies $u''(x)$ by the same constant, which implies that the value of the rate of change has no significance by itself.

Thus, we seek to find a measure based on the rate of change of the marginality of the utility function but modified so it remains unaltered before positive affine transformations. In that sense, we will define the Arrow-Pratt coefficient of absolute risk aversion.

Definition 2.10. *The **Arrow-Pratt measure of absolute risk aversion (ARA)** is defined as*

$$A(x) = -\frac{u''(x)}{u'(x)}$$

Under risk aversion, function A is positive. The absolute risk aversion directly measures the insistence of an individual for more-than-fair odds, at least when the bets are small. To measure a decision-maker's aversion to risk, it is natural to consider his risk premium for a small, actuarially neutral lottery. We are asking how much the agent is ready to pay to get rid of a zero-mean lottery that has a random pay-off of $\tilde{\epsilon}$. The answer to this question will be referred to as the risk premium ρ associated with that risk. If the lottery is accepted, the expected utility is given by $E(u(x+\tilde{\epsilon}))$. Then, the risk premium for the lottery (asset) $\tilde{\epsilon}$ is defined as

$$u(x-\rho) = E(u(x+\tilde{\epsilon})) \tag{2.2}$$

.

If $\tilde{\epsilon}$ has an expectation that differs from 0, we usually use the certainty equivalent, which as commented in def. 2.7, is the sure increase in wealth that has the same effect on welfare than bearing the risk. Since we know $E(\tilde{\epsilon}) = 0$, and by taking the Taylor expansion around $\rho = 0$ of the left hand side of eq. 2.2, we get

$$u(x-\rho) = u(x) - \rho u'(x) + O(\rho^2),$$

and by doing the same with the right hand side around $\tilde{\epsilon} = 0$,

$$E(u(x+\tilde{\epsilon})) = E(u(x) + \tilde{\epsilon}u'(x) + \frac{1}{2}\tilde{\epsilon}^2 u'(x) + O(\tilde{\epsilon}^3)) = u(x) + \frac{1}{2}\sigma_{\tilde{\epsilon}}^2 u''(x) + o(\sigma_{\tilde{\epsilon}}^2).$$

It follows that

$$\rho = \frac{1}{2}A(x)\sigma_{\tilde{\epsilon}}^2 + o(\sigma_{\tilde{\epsilon}}^2), \tag{2.3}$$

where

$$A(x) = -\frac{u''(x)}{u'(x)} = -\frac{d}{dx}\log u'(x)$$

Hence, $A(x)$ is the factor by which an economic agent with utility function u weighs the risk, measured by half the variance of $\tilde{\epsilon}$. That is, $A(x)$ is twice the risk premium per unit of variance for infinitesimal risks and, thus, variance might appear to be a good measure of the degree of riskiness of a lottery. An example of this is the use of a mean-variance decision criterion for modelling behaviour under risk. In a mean-variance model, we assume that individual risk attitudes depend only upon the mean and the variance of the underlying risks. However, the validity of these models is dependent on the degree of accuracy of the approximation of the risk premium, which can be considered accurate only when the

risk is small or in very special cases. Nonetheless, in most cases, the risk premium associated with any risk will also depend upon the other moments of the distribution of the risk, not just its mean and variance; e.g., two risks with the same mean and variance, but one with a distribution that is *skewed*[9] to the right and the other one skewed to the left, should not be expected to have the same risk premium.
Now suppose that two agents, with utility functions u and v respectively, have the same amount of money x, which is arbitrary. The agent v is more risk-averse than the agent u if *any* risk that is undesirable for agent u is also undesirable for agent v, i.e. the risk premium of any risk is larger for agent v than for agent u (Eeckhoudt, Gollier and Schlesinger [9]).

Proposition 2.1. *The following three condition are equivalent:*

i) The risk premium of any risk is larger for agent v than for agent u.

ii) For all x, $A_v(x) \geq A_u(x)$.

iii) $v(\cdot)$ is a concave transformation of function $u(\cdot)$, i.e. $\exists \phi(\cdot)$ with $\phi' > 0$ and $\phi'' \leq 0$ s.t. $v(x) = \phi(u(x))$ for all x.

Proof. We just need to prove that *i*) implies *ii*), *ii*) implies *iii*) and *iii*) implies *i*):

- *i*)$\Longrightarrow$*ii*). It is immediate from eq. 2.3.
- *ii*)$\Longrightarrow$*iii*). We have $A_v(x) \geq A_u(x)$. We say that v is more concave than u in the sense of Arrow-Pratt, and it is equivalent to the condition that v is a concave transformation of u: there exists an increasing and concave function ϕ s.t. $v(x) = \phi(u(x))$. Indeed, we have that

  $$v'(x) = \phi'(u(x))u'(x) \text{ and } v''(x) = \phi''(u(x))(u'(x))^2 + \phi'(u(x))u''(x)$$

  which implies that

  $$A_v(x) = A_u(x) + \frac{-\phi''(u(x))u'(x)}{\phi'(u(x))}$$

  Thus, A_v is uniformly larger than A_u if, and only if, ϕ is concave.

[9]The degree of skewness refers to the distortion or asymmetry of the probability distribution of a real-valued random variable. The skewness of a random variable X is the third standardised moment $\tilde{\mu}_3$, defined as

$$\tilde{\mu}_3 = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3}$$

where μ is the mean, σ the standard deviation and μ_3 is the third central moment.

- *iii)*$\Longrightarrow$*i)*. Consider a lottery μ and let ρ_v and ρ_u denote the risk premiums for the lottery μ of each agent. We have, by the definition of utility,

$$v(x-\rho_v) = E(v(x+\mu)) = E(\phi(u(x+\mu)))$$

Define the random variable λ as $\lambda = u(x+\mu)$. Since ϕ is concave, by Jensen's inequality: $E(\phi(\lambda)) < \phi(E(\lambda))$. It follows that

$$v(x-\rho_v) \leq \phi(E(\lambda)) = \phi(u(x-\rho_u)) = v(x-\rho_u)$$

And thus, since v is increasing, ρ_v is larger than ρ_u.

□

Risk aversion is driven by the fact that marginal utility is decreasing with wealth. But, how is the risk premium for an actuarially neutral risk, μ, affected by a change in the initial wealth? Intuitively, we can argue that richer people are less willing to pay to eliminate a fixed risk. For example, if we have a lottery for which it is equally probable to gain 100€ or none; it might be life-threatening for an individual with initial wealth of 101€, but it is essentially trivial for an agent with initial wealth of 1000000€. In that sense, we say that an agent with utility function u exhibits decreasing absolute risk aversion (resp. increasing) if it is less risk averse (resp. more risk averse) at higher levels of wealth. Decreasing or increasing absolute risk aversion are referred to jointly as monotone absolute risk aversion.

Definition 2.11. *For every $a \geq 0$, define a utility function u_a by $u_a(x) = u(x+a)$. Then u exhibits* ***decreasing absolute risk aversion*** *(resp. increasing) if, for all $a \geq 0$, u is more risk averse (resp. less) than u_a. It exhibits constant absolute risk aversion if it shows both increasing and decreasing absolute risk aversion.*

We can now characterise the monotonicity of the absolute risk aversion with the following theorem (Nielsen [25]).

Theorem 2.2. *Let u be a strictly increasing, risk averse utility function. The following three statements are equivalent:*

i) u exhibits decreasing (resp. increasing) absolute risk aversion.

ii) u is differentiable with $u' > 0$ and the cumulative risk aversion function $-\log u'(x)$ is concave (resp. convex).

iii) The index of ARA associated with u is decreasing (resp. increasing).

Proof. We will start by presenting a couple of lemmas which will be needed to prove the theorem

Lemma 2.5. *Let u be a strictly increasing function. Suppose that u exhibits decreasing or increasing absolute risk aversion. Then u is differentiable with $u' > 0$.*

Lemma 2.6. *Let u and v be two differentiable utility functions with positive first derivative. The u is more risk averse than v iff u'/v' is continuous and*

$$\log u'(r) - \log u'(s) \geq \log v'(r) - \log v'(s)$$

for $r, s \in S$, $r < s$.

Now we can give the proof for the theorem.

- *i)*$\Longrightarrow$*ii)*. We firstly observe that u' must be continuous. If there was a point x where u' was discontinuous, then $\log \circ u'$ would be discontinuous at x. From lemmas 2.5 and 2.6 we get that $\log u'(x) - \log u'(x+h)$ is continuous and decreasing (incresing) for all $h \geq 0$, then u' is discontinuous at $x+h$, for all $h \geq 0$. However, since u is risk averse, u' is decreasing, which implies that it is continuous almost everywhere. This is a contradiction, so u' is indeed continuous. Now, let $n \in \mathbb{N}$, set $h = 2^{-n}$ and $\mathcal{D}_n = \left\{\frac{a}{2^n} \in S \text{ s.t. } a \in \mathbb{Z}\right\}$. The function $\log u'(x) - \log u'(x+h)$ is decreasing on $\left\{x \in \mathcal{D}_n : x+h \in \mathcal{D}_n\right\}$. Therefore, the function $-\log \circ u'$ is concave (convex) on $\mathcal{D}_n$. It follows that it is concave on the set

  $$\mathcal{D} = \bigcup_{n=1}^{\infty} \mathcal{D}_n = \left\{\frac{a}{2^n} \in S \text{ s.t. } n \in \mathbb{N} \text{ and } a \in \mathbb{Z}\right\}$$

  Since $\mathcal{D}$ is dense in S, this function is concave (convex) on S.

- *ii)*$\Longrightarrow$*iii)*. Since $-\log \circ u'$ is concave (convex), it is absolutely continuous with the ARA index equal to its derivative. The function u has the ARA index $A(x) = (-\log \circ u')'$, which is decreasing (increasing).

- *iii)*$\Longrightarrow$*i)*. Since we have

  $$\log u'(x) - \log u'(x+h) = \int_x^{x+h} A(t)dt$$

  u is differentiable with $u' > 0$, and $A(x)$ is decreasing (increasing), the function $\log u'(x) - \log u'(x+h)$ is continuous and decreasing (increasing) for all $h \geq 0$.

□

Definition 2.12. *In general, the growth rate for a function $f(x)$ is defined as*

$$\frac{df(x)}{dx}\cdot\frac{1}{f(x)}=\frac{d\log f(x)}{dx}$$

Since marginal utility $u'(x)$ declines in wealth, its growth rate is negative. The absolute value of this negative growth rate, which is the measure of absolute risk aversion, is called the ***decay rate****.*

Absolute risk aversion is the rate of decay for marginal utility; it measures the rate at which marginal utility decreases when wealth is increased by one euro. But using different monetary units would change the ARA, i.e. the index of absolute risk aversion is not unit-free. Hence, it can be really useful, for a more accurate economic analysis, to use a unit-free measure. To this end, we can define the index of relative risk aversion as the rate at which utility decreases when wealth is increased by **one percent**.

Definition 2.13. *The measure of* ***relative risk aversion (RRA)*** *is defined as*

$$R(x)=-\frac{xu''(x)}{u'(x)}=xA(x)$$

Let $\tilde{\rho}(\mu)$ be the *relative risk premium* corresponding to a proportional risk μ; that is, a decision-maker with assets x and utility function u would be indifferent between receiving a risk $x\mu$ and receiving the non-random amount $E(x\mu)-x\tilde{\rho}(\mu)$. It is implicitly defined via the equation $u(x-x\tilde{\rho})=E(u(x+x\mu))$, and we get the following equality,

$$\tilde{\rho}(\mu)=\frac{1}{x}\rho(x\mu)$$

For a small, actuarially neutral, proportional risk μ, we have, by eq. 2.3,

$$\tilde{\rho}(\mu)=\frac{1}{2}\sigma_{\mu}^{2}R(x)+o(\sigma_{\mu}^{2}).$$

If μ is not actuarially neutral, we would have $\tilde{\rho}(\mu)=\frac{1}{2}\sigma_{\mu}^{2}R(x+xm(\mu))+o(\sigma_{\mu}^{2})$. The relative risk premium is also a unit-free measure, unlike the absolute risk premium. Obviously, if we normalise the initial wealth to unity, the relative and absolute risk premiums are equal.

To finish with this section, we will give two examples of classical utility functions. The first one is the so-called ***constant absolute risk aversion (CARA)***. These are exponential functions characterised by

$$u(x)=-\frac{e^{-ax}}{a}$$

where a is a positive scalar. Since $A(x) = (\log u')'(x)$, we get that $A(x) = a$ for all x. The fact that risk aversion is constant, is often useful in analysing choices among several alternatives. This assumption eliminates the income effect when dealing with decisions to be made about a risk whose size is invariant to changes in wealth.

The second example is the so-called ***constant relative risk aversion (CRRA)***, which is also called hyperbolic absolute risk aversion (HARA). The set of all CRRA utility functions is completely defined by

$$u(x) = \begin{cases} \frac{x^{1-\gamma}}{1-\gamma}, & \text{for } \gamma \geq 0, \gamma \neq 1, \\ \log(x), & \text{for } \gamma = 1 \end{cases}$$

This class of utility functions eliminates any income effects when making decisions about risks whose size is proportional to one's level of wealth. The assumption that relative risk aversion is constant simplifies many of the problems often encountered in macroeconomics and finance.

Chapter 3

Critique of the EU theory

Expected utility theory has been generally accepted as a normative model of rational choice. The theory suggests that people should act according to certain decision rules, but not that they will necessarily do so in reality. Traditional utility theory states that in decisions made without uncertainty, an individual should choose the alternative resulting in the highest level of utility, and decisions under uncertainty should be made according to the expected utility, maximising it. At first, it was also widely applied as a descriptive model of economic behaviour, however it has been shown that people infringe normative decision theory in many ways. The two most common examples of these violations are the *Allais paradox* and the *Ellsberg paradox*, but there are a few more. One of these is the so-called *framing effect* which was described by Tversky and Kahneman [36]. This effect consists in a violation of the *invariance* of decisions, which is described as follows: different representations of the same choice should yield the same preference, i.e., if two decision situations involve the same outcomes with the same probabilities resulting from the corresponding alternatives, then the person's decision should be identical for both. They also described the *endowment effect* and *loss aversion*: people's maximum willingness to pay to acquire an object is typically lower than the least amount they are willing to accept to give up the same object when they own it, even when there is no cause of attachment or if it was obtained minutes ago. In this chapter we will review these violations of the expected utility theory to understand how people's behaviour towards choice is, sometimes, not rational *according* to the theory that axiomatises it. Expected utility can be justified on the basis of a set of relatively simple axioms. The virtue of this approach is that it allows you to obtain a better understanding of what underlies the acceptance of the rule of maximising expected utility. If you accept all the axioms, then you are logically compelled to accept the maximisation of expected utility as the choice criterion, but if you reject one or more of the axioms, then EU does not necessarily follow. By studying how *reasonable* people respond to paradoxes and infringements of the axioms, we may have a better basis for deciding whether to accept the axioms and, hence, the EU theory.

3.1 The Framing effect

This is perhaps, the most fundamental critique, not merely of the EU theory, but of the much standard economic theory of optimal decision-making. People's perception and evaluation of outcomes is importantly affected by a reference point, which may be the person's *status quo*, or something to which his attention is drawn by the way the issue is **framed**. Following Tversky and Kahneman [36], the frame that a decision-maker adopts is controlled partly by the formulation of the problem and partly by the norms, habits, and personal characteristics of the DM. Rational choice requires that the preference between options should not reverse with changes of frame. Giving empirical evidence, they show that systematic reversals of preference are obtained when varying the framing of acts, contingencies, or outcomes. Let's show it with an example.

Example 3.1. *Two groups of people were given two surveys. In both surveys, respondents had to choose between two programmes to combat the outbreak of an unusual disease, which was expected to kill 600 people. The exact scientific estimate of the consequences of the programmes in the first survey was as follows:*

Programme A: 200 people will be saved.

Programme B: there is a 1/3 probability of saving 600 people, and 2/3 that no people will be saved.

In the second survey, given the same cover story of the first one, the programmes were presented as follows:

Programme C: 400 people will die.

Programme D: there is a 1/3 probability that nobody will die, and 2/3 that 600 people will die.

This test resulted in 72% of the people choosing programme A over B, and 78% of people choosing D over C, which is a contradiction in the sense of the EU theory. The expected value of all programmes is the same, but not the frame. The certain death of 400 people is less acceptable than the 2/3 chances of 600 people dying, and certainly saving 200 people is more attractive than a risky prospect of equal expected value. The preferences in both surveys illustrate a common pattern:

choices involving gains are often risk averse and choices involving losses are often risk-taking. They observed this pattern in several groups of respondents, which pointed in the direction of a framing effect and contradictory attitudes towards risk involving gains and losses.

3.2 The Endowment effect

As we introduced at the beginning of the chapter, the endowment effect is the finding that people are more likely to retain an object than acquire the same object when they do not own it. It was first described by Tehler [34], and then was demonstrated experimentally by Kahneman, Knetsch and Tehler [18]. They stated that this effect is a manifestation of loss aversion, the generalisation that losses are weighted substantially more than objectively commensurate gains in the evaluation of prospects and trades. To demonstrate the loss aversion relative to a *reference point*, they asked subjects to give a monetary value to a cheap decorative mug. The catch was that some subjects had been given the mug beforehand, but not all the participants. The mugs were assigned randomly, so there was no apparent reason that the underlying preferences should differ between those with mugs and those without. They found, however, that the subjects who had the mug prior to the valuation, gave a much higher value on it than those who didn't have it. In two separate experiments they found evidence of people tending to give much higher values on items they possessed, or were *endowed* with. In the first one, the median value for the mug among the subjects that had one was 7.12$, and 3.12$ among those who didn't have one. In the second one, the values were 7.00$ and 3.50$. In the theoretical models of the reference-dependent preferences (such as the work of Tversky and Kahneman [35]), the determination of the reference point around which losses and gains are encoded was left undetermined. It was taken to be the *status quo*, the current level of assets, or a level of aspiration or expectation. Model extensions have added discipline to this fact; we will introduce the model of Kőszegi and Rabin [19], which gives expectations-based mechanisms for the determination of stochastic reference distributions. It predicts that when risk is expected, and hence the referent is stochastic, behaviour will be different from when risk is unexpected and the referent is certain. In particular, when the referent is stochastic and the individual is offered a certain amount, the model predicts near risk neutrality. Contrarily, when the referent is a fixed certain amount and the individual is offered a gamble, it predicts risk aversion. Therefore, the Kőszegi and Rabin model features an **endowment effect for risk**. Let's build up the model. As commented above, the KR model is based on the Reference-Dependent model and Prospect theory of Tversky and Kahneman. A person's utility depends not only on his K-dimensional consumption bundle c, but also on a reference bundle r. He has an intrinsic consumption utility $m(c)$ that corresponds to the outcome-based utility. Overall utility is given by

$$u(c|r) \equiv m(c) + n(c|r),$$

where $n(c|r)$ is the so-called gain-loss utility. Both consumption utility and gain-loss utility are separable across dimensions, so that $m(c) \equiv \sum_k m_k(c_k)$ and $n(c|r) \equiv \sum_k n_k(c_k|r_k)$.[1] The person's gain-loss utility in dimension k depends solely and in a universal way on how consumption utility in that dimension compares to the consumption utility from the reference level. Hence, they assume that $n_k(c_k|r_k) \equiv \gamma(m_k(c_k) - m_k(r_k))$, where the function $\gamma(\cdot)$ satisfies the properties of the Tversky and Kahneman's value function:

1. $\gamma(x)$ is continuous for all x, twice differentiable for $x \neq 0$, and $\gamma(0) = 0$.
2. $\gamma(x)$ is strictly increasing.
3. If $y > x > 0$, then $\gamma(y) + \gamma(-y) < \gamma(x) + \gamma(-x)$.
4. $\gamma''(x) \leq 0$ for $x > 0$, and $\gamma''(x) \geq 0$ for $x < 0$.
5. $\frac{\gamma'_-(0)}{\gamma'_+(0)} \equiv \lambda > 1$, where $\gamma'_+(0) \equiv \lim_{x\to 0} \gamma'(|x|)$ and $\gamma'_-(0) \equiv \lim_{x\to 0} \gamma'(-|x|)$.

Loss aversion is captured by assumption 3 for large stakes and assumption 5 for small stakes. Assumption 4 captures another important feature of gain-loss utility, diminishing sensitivity: the marginal change in gain-loss sensations is greater for changes that are close to one's reference level than for changes that are further away. In their model, Kőszegi and Rabin specify a person's utility for a riskless outcome as $u(c|r)$, where $c = (c_1, c_2, \ldots, c_K) \in \mathbb{R}^K$ is consumption and

[1] We need three definitions before introducing the theorem for that assumption. A preference relation $\succeq$ has *jointly separable indices* if for any $E \subseteq I$ and all $x, y, z, z' \in X$, $x_E z \succeq y_E z \Longleftrightarrow x_E z' \succeq y_E z'$, where

$$x_E y := \begin{cases} x_i, & i \in E \\ y_i, & i \notin E \end{cases}$$

A preference relation $\succeq$ on $X = Z^{T+1}$ is *stationary* if for all $c \in Z, x, y \in Z^T$, $(c, x_0, \ldots, x_{T-1}) \succeq (c, y_0, \ldots, y_{T-1}) \Longleftrightarrow (x_0, \ldots, x_{T-1}, c) \succeq (y_0, \ldots, y_{T-1}, c)$. And we say that $\succeq$ is *sensitive* if all the indices are non-null (an index i is null if for all $x, y, z \in X$, $x_i z \sim y_i z$).
[Fishburn [11]]. A complete, transitive preference $\succeq$ on a set $X := \times_{t\in T} Z$ is continuous, stationary, sensitive, and has jointly separable indices iff there is a number $\delta > 0$ and a continuous non-constant function $u : Z \to \mathbb{R}$ s.t. $\succeq$ is represented by $U(z_0, z_1, \ldots) = \sum_t \delta^t u(z_t)$. Moreover, δ is unique and u is unique up to affine transformations.

$r = (r_1, r_2, \dots, r_k) \in \mathbb{R}^K$ is a *reference level* of consumption. If c is drawn according to the probability measure μ, the person's utility is given by

$$U(\mu|r) = \int u(c|r)d\mu(c)$$

But for them, the reference point itself is stochastic, so they suppose that the person's reference point is the probability measure η over $\mathbb{R}^K$, then the utility is as follows

$$U(\mu|\eta) = \int\int u(c|r)d\eta(r)d\mu(c)$$

with

$$u(c|r) = m(c) + \gamma(m(c) - m(r))$$

As they claim, this formulation captures the notion that the sense of gain or loss from a given consumption outcome derives from comparing it with all outcomes possible under the reference lottery. The important assumption of $m(c)$ and $n(c|r)$ being separable, together with loss aversion, is the essential point of many implications of reference-dependent utility, including the endowment effect. In evaluating an outcome, the DM assesses gain-loss utility in each dimension separately. This utility function replicates a number of properties commonly associated with reference-dependent preferences. In the following proposition, it is established that fixing the outcome, a lower reference point makes a person happier; and preferences exhibit a status quo bias, a preference for the current state of affairs.

Proposition 3.1. *If γ satisfies the properties 1-5, then the following hold.*

i) For all μ, η, η', such that for all $k \in \{1, \dots, K\}$, the marginal η'_k first-order stochastically dominates η_k, $U(\mu|\eta) \geq U(\mu|\eta')$.

ii) For any c, $c' \in \mathbb{R}^K$, $c \neq c'$, $u(c|c') \geq u(c'|c') \Longrightarrow (uc|c) > u(c'|c)$.

iii) Suppose γ satisfies that for all $x \neq 0$, $\gamma''(x) = 0$.[2] Then, for any μ, μ' such that $\mu \neq \mu'$, $U(\mu|\mu') \geq U(\mu'|\mu') \Longrightarrow U(\mu|\mu) > U(\mu'|\mu)$.

Points *ii*) and *iii*) mean that if a person is willing to abandon his reference point for an alternative, then he strictly prefers the alternative if that is his reference point. When the function $m(\cdot)$ is linear, the utility $u(c|r)$ exhibits the same properties as $\gamma(\cdot)$.

[2]When we are interested in characterising the implications of reference dependence where only loss aversion plays a role, we define an alternative assumption to assumption 4: for all $x \neq 0$, $\gamma''(x) = 0$.

Proposition 3.2. *If m is linear and γ satisfies properties 1-5, then there exists $\{\nu_k\}_{k\in K}$ satisfying the properties, such that for all c and r,*

$$u(c|r) - u(r|r) = \sum_{k\in K} \nu_k(c_k - r_k)$$

This equivalence does not hold when the changes are large or marginal consumption utilities change quickly. This is a good thing, as when large losses in consumption or wealth involved, diminishing marginal utility of wealth is likely to counteract the diminishing sensitivity in losses emphasised in prospect theory. Now consider a certain referent r, and a binary consumption gamble with outcomes $c_1 \geq r$ with probability α, and $c_2 \leq r$ with probability $1-\alpha$. The KR utility is

$$U(\mu|r) = \alpha u(c_1|r) + (1-\alpha)u(c_2|r)$$

If $c_1 \geq r > c_2$, the model predicts loss aversion to be present in the second term. The utility becomes

$$U(\mu|r) = \alpha[c_1 + 1\cdot(c_1 - r)] + (1-\alpha)[c_2 + \lambda(c_2 - r)]$$

Comparing this to the utility of the certain amount, $U(r|r) = r$, we have $U(\mu|r) > U(r|r)$, and the probability premium will be obtained for some $\widetilde{\mu}$, with corresponding probability $\widetilde{\alpha}$ s.t. $U(\widetilde{\mu}|r) = U(r|r)$:

$$\widetilde{\alpha} = \frac{r - c_2 - \lambda(c_2 - r)}{c_1 - c_2 + [(c_1 - r) - \lambda(c_2 - r)]}$$

Then, we can see a relationship between risk aversion elicited as $\widetilde{\alpha}$ and loss aversion, λ. For someone who is not loss averse, $\lambda = 1$, $\widetilde{\alpha} = \frac{r-c_2}{c_1-c_2}$. Then $r = \widetilde{\alpha}c_1 + (1-\widetilde{\alpha})c_2$. Risk neutral behaviour is exhibited by individuals who are not loss averse. For a loss averse individual, $\lambda > 1$ and $\widetilde{\alpha} > \frac{r-c_2}{c_1-c_2}$ for $c_1 > r > c_2 \geq 0$. The gamble $\widetilde{\mu}$ will have a higher expected value than r. Moreover, $\frac{d\widetilde{\alpha}}{d\lambda} > 0$ s.t. probability premiums are increasing in the degree of loss aversion. If endowed with a fixed amount in a probability premium task and trading for a gamble, a loss averse person will appear risk averse.

3.3 The Allais paradox

The Allais paradox is the best-known violation of the Independence axiom, and the most famous example of the supposed inconsistency of the EU theory. Maurice Allais [1] was one of the earliest researchers to demonstrate a deviation from the EU theory. He showed that people overweight outcomes that are considered certain, relative to outcomes which are merely probable. This phenomenon was labelled by Tversky and Kahneman [35] as the *certainty effect*. Allais distributed two surveys in which people were presented a pair of choice problems. We will follow the example described by Tversky and Kahneman, which differs from the original in that it refers to moderate rather than to extremely large gains. In the first survey, which we will denote as A, respondents were asked to choose between the following two lotteries:

$$A : \begin{cases} \mu_1 = 0.33\delta_{2500} + 0.66\delta_{2400} + 0.01\delta_0 \\ \lambda_1 = \delta_{2400} \end{cases}$$

The first lottery yields 2500€ with a probability of 0.33, 2400€ with a probability of 0.66 and draws a blank with the remaining probability. The second lottery yields 2400€ for sure. In the second survey (B), the respondents were presented with the following choice problem:

$$B : \begin{cases} \mu_2 = 0.33\delta_{2500} + 0.67\delta_0 \\ \lambda_2 = 0.34\delta_{2400} + 0.66\delta_0 \end{cases}$$

In survey A, most people chose the lottery λ_1, even though its expected value is lower than the one in μ_1, namely 2400€. The empirical test showed that 82% preferred the sure amount over the largest expected value. In survey B, the expected value of μ_2 is 825€ and for the lottery λ_2 it is 816€. The data showed that 83% of respondents preferred the slightly riskier lottery μ_2, in accordance with expectations. Moreover, the analysis of individual patterns of choice indicated that a majority of respondents made the modal choice in both problems, namely 61%. At least 65% of people chose both $\lambda_1 \succ \mu_1$ and $\mu_2 \succ \lambda_2$. This simultaneous choice leads to a *paradox* in the sense that it is inconsistent with the vNM paradigm. Note that B is obtained from A by eliminating the 0.66 chance of winning 2400€ from both prospects under consideration. Evidently, this change produces a greater reduction in desirability when it alters the character of the prospect from a sure gain to a probable one, than when both the original and the reduced prospects are uncertain. The experiment showed that at least 65% of the respondents violated the independence axiom, and therefore can be taken as empirical evidence against the vNM theory as a descriptive theory.

The main explanation for this paradox is the *certainty effect*. Reducing the probability of a negative outcome from 0.01 to 0 is usually judged to be a greater improvement than reducing the probability of that same outcome from, say, 0.34 to 0.33. But that is a psychological approach to the problem; if we want to give a mathematical explanation to the paradox, we can use *subjective distortions* of objective lotteries.

As Allais claims in Allais and Hagen [2], the subjective distortion of objective probabilities generally seems to depend on whether gain or loss is at issue, as well as the amounts involved. The same person is perfectly capable of cautiously buying insurance protection against fire, and losing all his fortune at the races. Whereas the gambler overestimates the probability of gain and underestimates the likelihood of loss, the prudent people tend to overestimate the probability of loss and underestimate the probability of gain. Take $\mathcal{X}$ as a set of bounded measurable functions X on some measurable set $(\Omega, \mathcal{F})$. In this approach, which was presented by Leonard Savage in Foundations of Statistics [31], probabilities are not known in advance, so we work with uncertainty instead of risk. Following Föllmer and Schied [14], assuming that $\mathcal{X}$ is endowed with a preference relation $\succ$. Savage gave seven postulates (completeness and transitivity, monotonicity, sure-thing principle, absence of indifference, continuity and two independence axioms) to guarantee that there existed a numerical representation of the special form

$$U(X) = E_Q[u(x)] = \int u(X(\omega))Q(d\omega) \text{ for all } X \in \mathcal{X}$$

where Q is a probability measure on $(\Omega, \mathcal{F})$ and u is a function on $\mathbb{R}$. It is the measure Q that specifies the subjective view of the probabilities of events. This subjectivity is implicit in the preference relation. Since $\succ$ is *monotone* in the sense that

$$Y \succeq X \text{ if } Y(\omega) \geq X(\omega) \text{ for all } \omega \in \Omega$$

is equivalent to the condition that u is an increasing function. Let P be an objective probability measure on $(\Omega, \mathcal{F})$, then the preference relation $\succ$ may be such that the subjective measure Q is different from P. Suppose that P is a Lebesgue measure restricted to $\Omega = [0, 1]$, and that $\mathcal{X}$ is the space of bounded right continuous increasing functions on $[0, 1]$.

Definition 3.1. *A function $f : (c, d) \to (a, b)$ is called an inverse function for an increasing function $F : (a, b) \to \mathbb{R}$ if*

$$F(f(s)-) \leq s \leq F(f(s)+) \textit{ for all } s \in (c, d)$$

Lemma 3.1. *Let Y be a random variable on a probability space $(\Omega, \mathcal{F}, P)$ with a uniform distribution on (0,1). If f is an inverse function of a normalised increased right-continuous function $F : \mathbb{R} \to [0, 1]$, then*

$$X(\omega) := f(Y(\omega))$$

has the distribution function F.

Proof. Any inverse function for F is measurable because it coincides with the measurable function f^+ outside the countable set $\{s \in (0,1) \mid f^-(s) < f^+(s)\}$.[3] To continue with the proof, we need the following lemma

Lemma 3.2. *Let f be an inverse function for F. Then F is an inverse function for f and, in particular,*

$$F(x+) = \inf\{s \in (c,d) \mid f(s) > x\} \text{ for } x \text{ with } F(x) < d,\ c < d.$$

Since $f(F(x)-) \leq x$, we have $f(s) \leq x$ for $s < F(x)$, and by the lemma, $f(s) \leq x$ implies $F(x) \geq F(f(s)) = F(f(s)+) \geq s$. It follows that

$$(0, F(x)) \subseteq \{s \in (0,1) \mid f(s) \leq x\} \subseteq (0, F(x)]$$

and, therefore

$$F(x) = P[Y \in (0, F(x))] \leq P\big[Y \in \{s \mid f(s) \leq x\}\big] \leq P\big[Y \in (0, F(x)]\big] = F(x)$$

The assertion follows from the identity $P\big[Y \in \{s \mid f(s) \leq x\}\big] = P\big[X \leq x\big]$. □

Let us denote now $\mu_{P,X}$ as the distribution of X under P. By the lemma above, every probability measure on $\mathbb{R}$ with bounded support[4] is of the form $\mu_{P,X}$. If the agent agrees that, objectively, $X \in \mathcal{X}$ can be identified with a lottery $\mu_{P,X}$, and having the numerical representation $U^*(\mu_{P,X})$ of the preference order, then it may violate the independence axiom. The agent might take a pessimistic view and distort P by putting more emphasis on unfavourable scenarios, e.g., the agent could replace P by the subjective measure $Q = \alpha\delta_0 + (1-\alpha)P$. Replacing P by Q corresponds to a non-linear distortion on the level of lotteries, i.e. $\mu = \mu_{P,X}$ is distorted to $\mu^* = \mu_{Q,X}$ given by $\mu^* = \alpha\delta_{l(\mu)} + (1-\alpha)\mu$, where $l(\mu) := \sup\{a \in \mathbb{R} \mid \mu\big((-\infty, a)\big) = 0\}$. Now, going back to the Allais paradox, taking

$$\begin{aligned} \mu_1^* &= \alpha\delta_0 + (1-\alpha)\mu_1 \quad \text{and} \quad \lambda_1^* = \lambda_1 \\ \mu_2^* &= \alpha\delta_0 + (1-\alpha)\mu_2 \quad \text{and} \quad \lambda_2^* = \alpha\delta_0 + (1-\alpha)\lambda_2 \end{aligned}$$

Denoting $U^*(\mu) = \int u(x) d\mu^*(x)$, for the particular choice $u(x) = x$ we have $U^*(\mu_2) > U^*(\lambda_2)$ and for $\alpha > 9/2409$ we get $U^*(\lambda_1) > U^*(\mu_1)$, in accordance with what we have seen in the example of the paradox. Thus, we have seen that the Savage approach to utility theory is able to give an answer to why people act in an irrational way with respect to the classical EU theory.

[3] The functions $f^-(s)$ and $f^+(s)$ are called the left- and right-continuous inverse functions of F, and are defined as

$$f^-(s) := \inf\{x \in \mathbb{R} \mid F(x) \geq s\} \text{ and } f^+(s) := \sup\{x \in \mathbb{R} \mid F(x) \leq s\}$$

These are also known as *quantile functions*.

[4] The support of a Borel probability measure μ on $\mathbb{R}^d$ is the smallest closed set $A \subset \mathbb{R}^d$ such that $\mu(A^c) = 0$.

3.4 The Ellsberg paradox

The Allais paradox derives its force from our tendency to prefer having a good for certain to having a chance at a greater good. The Ellsberg paradox appeals to put preference for known risks over unknown ones (Segal [32]). Consider the following problem:
There are two urns denoted as X and Y. Urn X has 50 red balls and 50 green balls. Urn Y has 100 total balls, some red and the rest green, but the numbers of each are unknown. A person is offered a choice between two lotteries:

A: A ball is drawn at random from X; you get 10€ if red, 0 if green.

B: A ball is drawn at random from Y; you get 10€ if red, 0 if green.

Consider now the next choice problem:

C: A ball is drawn at random from X; you get 10€ if green, 0 if red.

D: A ball is drawn at random from Y; you get 10€ if green, 0 if red.

Turns out that many people faced with the first choice problem, preferred A. When faced with the second choice problem, many people preferred lottery C. More remarkably, when offered both choices in different questions, many people chose A in the first question and C in the second one. This raises a more basic issue about their behaviour as such choice is simply inconsistent with any subjective beliefs about the composition of red and green balls in urn Y. Since the composition of urn Y is unknown, there are no objective probabilities. Suppose that your subjective probability of drawing a red ball from Y is p, and hence that of drawing a green ball is $1-p$. Since the two lotteries in each comparison have the same prizes, a rational person who prefers more to less should prefer the lottery that offers a higher probability of winning the prize. This is irrespective of whether the preferences can be represented by expected utility; it is simply a matter of monotonicity. A person should prefer A to B if $p < 1/2$, and should prefer C to D if $1 - p < 1/2$, that is $p > 1/2$, which is a contradiction. There is no probability measure supporting these preferences through expected utility maximisation. The behaviour suggests ambiguity aversion; people dislike the added uncertainty about the risk: people seem to prefer *known unknowns* to *unknowns unknowns.* Now, getting back to the Savage model presented in the previous section, we can make one further conceptual step so the Ellsberg paradox can fit into the setting. Let's consider a class of measures $\mathcal{Q}$ on $(\Omega, \mathcal{F})$ instead of a single measure Q. The aim of the extension is to characterise those preference relations on $\mathcal{X}$ that admit a representation of the form

$$U(x) = \inf_{Q \in \mathcal{Q}} E_Q[u(X)]$$

We are going to embed $\mathcal{X}$ into a certain space $\widetilde{\mathcal{X}}$ of functions $\widetilde{X}$ with values in the convex set $\mathcal{M}_b(\mathbb{R}) = \{\mu \in \mathcal{M}_1(\mathbb{R}) \mid \mu([-c,c]) = 1 \text{ for some } c \geq 0\}$. $\widetilde{\mathcal{X}}$ is defined as the convex set of all those stochastic kernels[5] $\widetilde{\mathcal{X}}(\omega, dy)$ from $(\Omega, \mathcal{F})$ to $\mathbb{R}$ for which there exists a constant $c \geq 0$ s.t. $\widetilde{\mathcal{X}}(\omega, [-c,c]) = 1$, for all $\omega \in \Omega$. By the mapping

$$X \in \mathcal{X} \mapsto \delta_X \in \widetilde{\mathcal{X}}$$

the space $\mathcal{X}$ can be embedded into $\widetilde{\mathcal{X}}$. Therefore, a preference order on $\mathcal{X}$ with a representation as the one described above, extends to $\widetilde{\mathcal{X}}$ by

$$\widetilde{U}(\widetilde{\mathcal{X}}) = \inf_{Q\in\mathcal{Q}} \int \int u(y)\widetilde{X}(\omega, dy)Q(d\omega) = \inf_{Q\in\mathcal{Q}} E_Q[\widetilde{u}(\widetilde{X})]$$

where $\widetilde{u}$ is the affine function on $\mathcal{M}_b(\mathbb{R})$ defined by

$$\widetilde{u}(\mu) = \int u d\mu, \quad \mu \in \mathcal{M}_b(\mathbb{R})$$

Going back to the Ellsberg paradox. Set $\Omega = \{0,1\}$ and define

$$\widetilde{X}_0(\omega) := p\delta_{10} + (1-p)\delta_0, \quad \widetilde{X}_1 := (1-p)\delta_{10} + p\delta_0$$

and

$$\widetilde{Z}_i(\omega) := \delta_{10}\mathbb{1}_{\{i\}}(\omega) + \delta_0\mathbb{1}_{\{1-i\}}(\omega)$$

Now take $\mathcal{Q} := \{q\delta_1 + (1-q)\delta_0 \mid a \leq q \leq b\}$ with $[a,b] \subset [0,1]$. The functional

$$\widetilde{U}(\widetilde{X}) := \inf_{Q\in\mathcal{Q}} E_Q[\widetilde{u}(\widetilde{X})]$$

satisfies $\widetilde{U}(\widetilde{X}_i) > \widetilde{U}(\widetilde{Z}_i)$, $i = 0,1$, as soon as $a < p < b$, in accordance with the preferences described in the paradox.

[5]Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces. A stochastic kernel, or Markov kernel, is a map $\kappa : \mathcal{B} \times X \to [0,1]$ with the following properties:

i) For every $B \in \mathcal{B}$, the map $x \mapsto \kappa(B,x)$ is $\mathcal{A}$-measurable.

ii) For every $x \in X$, the map $B \mapsto \kappa(B,x)$ is a probability measure on $(Y, \mathcal{B})$.

Chapter 4

Monetary risk measures

A monetary risk measure is a mathematical tool for quantifying the risk of a random future gain, or loss, which is denoted in discounted units of a reference instrument, e.g. a currency (Hamel [15]). In chapter 3 we have given an extension to our model so it can explain the main paradoxes encountered in the theory. We ended the chapter by introducing a new utility function that is able to solve the contradiction in the preferences of the Ellsberg paradox This numerical representation corresponds to the so-called *robust preferences*. In fact, the numerical representation is defined as

$$U(x) = \inf_{Q \in \mathcal{Q}} \big(E_Q[u(X)] + \gamma(Q) \big)$$

The agent considers a whole class of probabilistic models specified by probability measures Q on the given set of scenarios, but different models Q are taken more or less seriously, and this is made precise in terms of the penalty function $\gamma(Q)$. In evaluating a given financial position, the agent then takes a worst case approach by taking the infimum of expected utilities over the suitably penalised models. Besides the monotonicity of the preferences that we presented in the previous chapter, we need to assume the following three axioms. The first two are suitable extensions of the two main axioms of vNM theory.

Axiom 4.1 (Weak certainty independence). *If for $\widetilde{X}, \widetilde{Y} \in \widetilde{\mathcal{X}}$, and for some $\nu \in \mathcal{M}_b(S)$ and $\alpha \in (0,1]$, we have $\alpha\widetilde{X} + (1-\alpha)\nu \succ \alpha\widetilde{Y} + (1-\alpha)\nu$, then*

$$\alpha\widetilde{X} + (1-\alpha)\mu \succ \alpha\widetilde{Y} + (1-\alpha)\mu \quad \textit{for all } \mu \in \mathcal{M}_b(S).$$

Axiom 4.2 (Archimedean axiom). *If $\widetilde{X}, \widetilde{Y}, \widetilde{Z} \in \widetilde{\mathcal{X}}$ are such that $\widetilde{Z} \succeq \widetilde{Y} \succeq \widetilde{X}$, then there are α, $\beta \in (0,1)$ with*

$$\alpha\widetilde{Z} + (1-\alpha)\widetilde{X} \succ \widetilde{Y} \succ \beta\widetilde{Z} + (1-\beta)\widetilde{X}.$$

Axiom 4.3 (Uncertainty aversion). *If $\widetilde{X}, \widetilde{Y} \in \widetilde{\mathcal{X}}$ are such that $\widetilde{X} \sim \widetilde{Y}$, then*

$$\alpha\widetilde{X} + (1-\alpha)\widetilde{Y} \succeq \widetilde{X} \quad \textit{for all } \alpha \in [0,1]$$

It is necessary to formulate the axiom of uncertainty aversion, but even without its axiomatic foundation, the representation of preferences in the face of model uncertainty by a subjective utility assessment is highly plausible as it stands. The agent penalises every possible probabilistic view $Q \in \mathcal{Q}$ in terms of the penalty $\gamma(Q)$ and takes the worst case approach in evaluating the pay-off of a given financial position. Removing this penalisation reduces the complexity of the mathematics required, and it is often referred to as the *coherent case*. Now, this representation of preferences, characterised in terms of a robust extension of the vNM axioms, can be reduced to the robust representation of convex risk measures.

Definition 4.1. *A financial position is a function $X : \Omega \to \mathbb{R}$. $X(\omega)$ is interpreted as the discounted net worth of the position scenario ω at the end of the period under consideration.*

Definition 4.2. *A risk measure is a mapping $\rho : \mathcal{X} \to \mathbb{R}$.*

Definition 4.3. *Given a subset $\mathcal{A} \subset \mathcal{X}$, we define the risk of a position $X \in \mathcal{A}$ by*

$$\rho_{\mathcal{A}}(X) = \inf\{m \in \mathbb{R} \ : \ X + m \in \mathcal{A}\}$$

With the convention of $\inf \emptyset = \infty$.

Definition 4.4. *Given a risk measure $\rho : \mathcal{X} \to \mathbb{R}$, we define the acceptance set of ρ by*

$$\mathcal{A}_\rho = \{X \in \mathcal{X} \ : \ \rho(X) \leq 0\}$$

Proposition 4.1. *Let $\mathcal{A} \subset \mathcal{X}$ be an acceptance set and $\rho_{\mathcal{A}}$ a risk measure, then the following properties are satisfied*

1. $\rho_{\mathcal{A}} > -\infty$.
2. *For each $m \in \mathbb{R}$, $\rho_{\mathcal{A}}(m)$ is finite.*
3. *$\rho_{\mathcal{A}}$ is monotone: $Y \geq X \Longrightarrow \rho_{\mathcal{A}}(Y) \leq \rho_{\mathcal{A}}(X)$.*
4. *$\rho_{\mathcal{A}}$ is cash-invariant, i.e., for all positions X and constants m, we have*

$$\rho_{\mathcal{A}}(X+m) = \rho_{\mathcal{A}} - m$$

5. *$\mathcal{A}$ is normalised if, and only if, $\rho_{\mathcal{A}}(0) = 0$.*

In terms of monetary risk, what really matters is the downside risk, which is an estimation of a security's potential to suffer a devaluation if the market conditions change, or the amount of loss that could be sustained as a result of the decline. Downside risk explains a worst case scenario for an investment or indicates how much the investor stands to lose. In that sense, the monotonicity is clear, the downside risk of a position is reduced if the pay-off profile is increased. Cash invariance is motivated by the interpretation of $\rho_{\mathcal{A}}$ as a capital requirement. Thus, if the amount m is added to the position and invested in a risk-free manner, the capital requirement is reduced by the same amount. In particular, cash invariance implies $\rho_{\mathcal{A}}(X + \rho_{\mathcal{A}}(X)) = 0$, i.e., the accumulate position consisting of X and the risk-free investment $\rho_{\mathcal{A}}$ is acceptable.

Example 4.1. *(**Worst case risk measure**) Extremely pessimistic people would choose the worst case risk measure, defined by*

$$\rho_{max}(X) := -\inf_{\omega\in\Omega} X(\omega)$$

which induces the acceptance set $\mathcal{A}_\rho = \{X \in \mathcal{X} \mid X(\omega) \geq 0 \text{ for all } \omega \in \Omega\}$. This risk measure is the least upper bound for the potential loss which can occur in any scenario. Thus, it is the most conservative (normalised) risk measure, in the sense that for any risk measure $\widetilde{\rho}$, $\widetilde{\rho} \leq \rho_{max}$.

Example 4.2. *(**Value at Risk**) Value at Risk estimates how much a set of investments might lose given normal market conditions (Chun, Shapiro and Uryasev [6]). The idea behind the measure is that very unlikely events should be neglected. Given a number $\alpha \in [0, 1)$, we regard a financial position as acceptable if the probability to obtain a negative value is less or equal to α. We assume that Ω is equipped with a σ -field $\mathcal{F}$ and a probability measure $\mathbb{P}$, and all elements in $\mathcal{X}$ are measurable with respect to $\mathcal{F}$. We choose $\mathcal{A} := \{X \in \mathcal{X} \mid \mathbb{P}[X < 0] \leq \alpha\}$. The corresponding monetary risk measure is called value at risk at level α, and denoted VaR_α:*

$$\begin{aligned} VaR_\alpha(X) &= \inf\{m \in \mathbb{R} \ : \ X + m \in \mathcal{A}\} \\ &= \inf\{m \in \mathbb{R} \ : \ \mathbb{P}[X < -m] \leq \alpha\} \end{aligned}$$

Value at Risk has been the standard risk measure because it is easy to understand and easy to calculate, as soon as $\mathbb{P}$ is specified. However, it can penalise diversification, which means taking a convex combination of different risky financial positions, i.e., if $X_1, X_2 \in \mathcal{X}$ are two different financial positions, then $X = \lambda X_1 + (1 - \lambda)X_2$ for some $\lambda \in [0, 1]$ is a diversification. The risk of X should be at most as high as the maximum risk of X_1 and X_2, it should not increase the risk. If two positions are acceptable, then each convex combination of them should be acceptable as well, that is the acceptance set should be convex.

Example 4.3. *Let* $\alpha = 0.01$*, and let* X_1*,*X_2 *be two independent random variables with* $\mathbb{P}[X_i = 1000] = 0.992$ *and* $\mathbb{P}[X_i = -10000] = 0.008$*. We have* $E(X_i) = 992 - 80 = 912$*, so that the expected return of each position equals* 9.12%*. The probability of a loss is less than* α *for each position, so that both of them are "acceptable".* $\sigma^2_{0.01}(X_i) = -1000$ *for both positions. It should be less risky to diversify. However the probability that at least one of the two variables is negative equals* $2 \cdot 0.008 - 0.008^2 \approx 0.016 > \alpha$*. Hence* X *is not "acceptable". Therefore, Value at Risk penalises diversification.*

To prevent this from happening, we can set some more axioms to guarantee that any risk measure will work properly in every situation.

4.1 Convex and coherent risk measures

Axiom 4.4 (**Convexity**)**.** *A monetary risk measure* ρ *is called convex risk measure if it satisfies*

$$\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y) \quad \textit{for all } \lambda \in [0,1]$$

A convex risk measure is called a **coherent risk measure** if it satisfies the following two axioms (Artzner [4]).

Axiom 4.5 (**Positive homogeneity**)**.** *For all* $\lambda \geq 0$*, and all* $X \in \mathcal{X}$*,*

$$\rho(\lambda X) = \lambda\rho(x)$$

Under the assumption of positive homogeneity, the convexity of a monetary risk measure is equivalent to sub-additivity.

Axiom 4.6 (**Sub-additivity**)**.** *For all* $X, Y \in \mathcal{X}$*,*

$$\rho(X+Y) \leq \rho(x) + \rho(Y)$$

The convexity axiom gives a precise meaning to the idea that diversification should not increase the risk. But this idea becomes even clearer when we note that, for a monetary risk measure, convexity is in fact equivalent to the weaker requirement of ***Quasi convexity***: $\rho(\lambda X + (1-\lambda)Y) \leq \max(\rho(X), \rho(Y))$, for $\lambda \in [0,1]$. If position size directly influences risk, e.g, if positions are large enough that the time required to liquidate them depends on their sizes, then we should consider the consequences of lack of liquidity when computing the future net worth of a position. With this in mind, positive homogeneity and sub-additivity remain reasonable. Positive homogeneity is imposed to model what a government or an

exchange might impose in a situation where no netting[1] or diversification occurs, in particular because the government does not prevent many firms from all taking the same position.
If we get back to the *Value at Risk* measure, we can see that it is a positively homogeneous monetary risk measure, but it is not coherent since the sub-additivity axiom is not satisfied. Instead, we can define the following risk measure.

Example 4.4. ***(Average Value at Risk)*** *Average Value at Risk at a level $\lambda \in (0,1]$ is defined as*

$$AVaR_\lambda = -\frac{1}{\lambda}\int_0^\lambda VaR_\alpha(X)d\alpha$$

It is also called ***Conditional Value at Risk*** *or* ***Expected Shortfall****.*

This measure estimates the risk of an investment in a conservative way, focusing on the less possible outcomes. For high values of λ it ignores the most profitable but unlikely possibilities, while for small values of λ it focuses on the worst losses.

Proposition 4.2. *If $\mathcal{A}_\rho$ is convex, then $\rho_\mathcal{A}$ is convex. If $\mathcal{A}_\rho$ is a cone (i.e., if for all $X \in \mathcal{A}$ and all $\lambda \geq 0$ we have $\lambda X \in \mathcal{A}$), then $\rho_\mathcal{A}$ is positively homogeneous and, in particular, if $\mathcal{A}_\rho$ is a convex cone, $\rho_\mathcal{A}$ is a coherent measure of risk.*

Proof. Suppose that $X_1, X_2 \in \mathcal{X}$ and that $m_1, m_2 \in \mathbb{R}$ are such that $m_i + X_i \in \mathcal{A}_\rho$. If $\lambda \in [0,1]$, then the convexity of $\mathcal{A}$ implies that $\lambda(m_1 + X_1) + (1-\lambda)(m_2 + X_2) \in \mathcal{A}_\rho$. Thus, by the cash invariance of $\rho_\mathcal{A}$,

$$\begin{aligned} 0 &\geq \rho_\mathcal{A}\big(\lambda(m_1 + X_1) + (1-\lambda)(m_2 + X_2)\big) \\ &= \rho_\mathcal{A}\big(\lambda X_1 + (1-\lambda)X_2\big) - \big(\lambda m_1 + (1-\lambda)m_2\big), \end{aligned}$$

and the convexity of $\rho_\mathcal{A}$ follows. For the second part of the proposition, as in the proof of convexity, we obtain that $\rho_\mathcal{A}(\lambda X) \leq \lambda \rho_\mathcal{A}(X)$ for $\lambda \geq 0$ if $\mathcal{A}_\rho$ is a cone. To prove the converse inequality, let $m < \rho_\mathcal{A}(X)$. Then $m + X \notin \mathcal{A}_\rho$ and hence $\lambda m + \lambda X \notin \mathcal{A}_\rho$ for $\lambda \geq 0$. Thus, $\lambda m < \rho_\mathcal{A}(\lambda X)$.

□

Definition 4.5. *If ρ is a convex risk measure, then $\phi(X) := -\rho(X)$ is called a concave monetary utility functional. If ρ is coherent, ϕ is called coherent monetary utility functional.*

[1]Netting entails offsetting the value of multiple positions or payments due to be exchanged between two or more parties. It can be used to determine which party is owed remuneration in a multiparty agreement.

By reversing the sign of the risk measure, we put emphasis on the utility of a position rather than on its risk.

Example 4.5. *(**Acceptance in terms of expected utility**) Let Ω be equipped with a σ-field $\mathcal{F}$ and a probability measure $\mathbb{P}$. Assume that all elements of $\mathcal{X}$ are measurable with respect to $\mathcal{F}$ and bounded. Let $u : \mathbb{R} \to \mathbb{R}$ be a strictly increasing and concave utility functions. Fix $x_0 \in \mathbb{R}$, and let*

$$\mathcal{A} := \{X \in \mathcal{X} \mid E[u(X)] \geq u(x_0)\}.$$

That is, X is accepted if, and only if, the certainty equivalent of X, i.e., $u^{-1}(E[u(X)])$, is greater than or equal to x_0. Moreover, concavity of u implies that for $X, Y \in \mathcal{A}$ and $\lambda \in [0,1]$,

$$\begin{aligned} E[u(\lambda X + (1-\lambda)Y)] &\geq E[\lambda u(X) + (1-\lambda)u(Y)] \\ &= \lambda E[u(X)] + (1-\lambda)E[u(Y)] \\ &\geq u(x_0) \end{aligned}$$

so that $\mathcal{A}$ is convex. The corresponding convex risk measure is given by

$$\rho_{\mathcal{A}}(X) = \inf\{m \in \mathbb{R} \mid E[u(X+m) \geq u(x_0)]\}.$$

Let us consider now the special case of the exponential utility function $u(x) = 1 - e^{-\beta x}$ with $\beta > 0$, a type of CARA utility function, and $x_0 = 0$. We have

$$\begin{aligned} u'(x) &= \beta e^{-\beta x} > 0 \\ u''(x) &= -\beta^2 e^{-\beta x} < 0 \end{aligned}$$

u is strictly increasing and concave. The condition $E[u(X+m)] \geq u(x_0)$ now reads as $1 - e^{-\beta m} E[e^{-\beta X}] \geq 0$ or, equivalently, $m \geq \frac{1}{\beta} \log E[e^{-\beta X}]$. Hence

$$\rho_{\mathcal{A}}(X) = \frac{1}{\beta} \log E[e^{-\beta X}],$$

which is called the **entropic risk measure**. This measure is a possible alternative to the Value at Risk and Average Value at Risk measures, and it is the typical example of a convex risk measure which is not coherent. The entropic risk measure can also be represented as

$$\frac{1}{\beta} \log E[e^{-\beta X}] = \sup_{Q \in \mathcal{M}_1(\Omega, \mathcal{F})} \left(E_Q[-X] - \frac{1}{\beta} H(Q|\mathbb{P}) \right)$$

where $H(Q|\mathbb{P}) = E\left[\frac{dQ}{d\mathbb{P}} \log \frac{dQ}{d\mathbb{P}}\right]$ is the relative entropy of $Q \ll \mathbb{P}$ (Q is absolutely continuous with respect to $\mathbb{P}$).

4.2 Dual representation

Suppose that $\mathcal{X}$ consists of measurable functions on $(\Omega, \mathcal{F})$. Following Föllmer and Schied [13], a *dual representation* of a convex risk measure ρ has the form

$$\rho(X) = \sup_{Q \in \mathcal{M}} \big(E_Q[-X] - \gamma(Q)\big).$$

Where $\gamma : \mathcal{M} \to \mathbb{R} \cup \{+\infty\}$ is the penalty function that we introduced at the beginning of the chapter. The elements of $\mathcal{M}$ can be interpreted as possible probabilistic models, which are taken more or less seriously according to the size of the penalty function. The value $\rho(X)$ is computed as the worst case expectation over all models $Q \in \mathcal{M}$ and penalised by $\gamma(Q)$. For these reasons, a representation of that form is also called a *robust representation*.

Definition 4.6. *For every $Q \in \mathcal{M}$, we define the minimal penalty function of ρ by*

$$\gamma_\rho(Q) := \sup_{X \in \mathcal{X}} \big(E_Q[-X] - \rho(X)\big) = \sup_{X \in \mathcal{A}_\rho} E_Q[-X]$$

Proposition 4.3. *For a convex risk measure ρ admitting a robust representation, the following statements are equivalent*

i) ρ is coherent.

ii) There exists a robust representation whose penalty function γ only takes the values 0 and ∞.

iii) γ_ρ only takes the values 0 and ∞.

Proof. The fact that the third statement implies the second one is straightforward. If γ only takes the values 0 and ∞, then ρ is positively homogeneous and, hence, coherent. Finally, to see that the first statement implies the third one, let $Q \in \mathcal{M}$ and $\lambda > 0$, and recall that the acceptance set of a coherent risk measure is a cone. Consequently,

$$\gamma_\rho(Q) = \sup_{X \in \mathcal{A}_\rho} E_Q[-X] = \sup_{X \in \mathcal{A}_\rho} E_Q[-\lambda X] = \lambda \gamma_\rho(Q).$$

Hence, γ_ρ can only take the values 0 and ∞. $\square$

A dual representation in terms of probability measures is closely related to certain continuity properties of ρ. A convex risk measure ρ which admits a robust representation on $\mathcal{M}_1$ is continuous from above in the sense that

$$X_n \searrow X \Longrightarrow \rho(X_n) \nearrow \rho(X)$$

Lemma 4.1 (Fatou property). *Continuity from above is equivalent to the so-called Fatou property: for any bounded sequence (X_n) converging pointwise to X*

$$\liminf_{n\to\infty} \rho(X_n) \geq \rho(X)$$

Proof. We first show that Fatou property holds. Dominated convergence implies that $E_Q[X_n] \to E_Q[X]$ for each $Q \in \mathcal{M}_1$. Hence,

$$\begin{aligned}\rho(X) &= \sup_{Q\in\mathcal{M}_1} \big(\lim_{n\to\infty} E_Q[-X_n] - \gamma(Q)\big)\\ &\leq \liminf_{n\to\infty} \sup_{Q\in\mathcal{M}_1} \big(E_Q[-X_n] - \gamma(Q)\big)\\ &= \liminf_{n\to\infty} \rho(X_n)\end{aligned}$$

In order to show the equivalence between continuity from above and the Fatou property we will first assume the latter. By monotonicity, $\rho(X_n) \leq \rho(X)$ for each n if $X_n \nearrow X$, and so $\rho(X_n) \searrow \rho(X)$ follows. Now, assuming continuity from above. Let (X_n) be a bounded sequence in $\mathcal{X}$ which converges pointwise to X. Define $Y_m := \sup_{n\geq m} X_n \in \mathcal{X}$. Then Y_m decreases $\mathbb{P}$-almost surely to X. Since $\rho(X_n) \geq \rho(Y_n)$ by monotonicity, continuity from above yields that

$$\liminf_{n\to\infty} \rho(X_n) \geq \lim_{n\to\infty} \rho(Y_n) = \rho(X)$$

□

4.3 Law-invariant risk measures

Here we discuss those convex risk measures ρ on $\mathcal{X} = \mathcal{L}^\infty(\Omega, \mathcal{F}, \mathbb{P})$[2]. Assuming that Ω is equipped with a σ-field $\mathcal{F}$ and a probability measure $\mathbb{P}$ and that all elements of $\mathcal{X}$ are measurable with respect to $\mathcal{F}$. If we regard $\mathbb{P}$ as the "true probability measure", it is natural to concentrate on law-invariant risk measures. Following Kusuoka [20], we will end this section with the Kusuoka's representation theorem.

Definition 4.7. *A monetary risk measure ρ is called law-invariant if $\rho(X) = \rho(Y)$ whenever X and Y have the same distribution under $\mathbb{P}$.*

If ρ is law-invariant, it will turn out that the only relevant aspect of the probability measures Q is the distribution of the density $dQ/d\mathbb{P}$ under $\mathbb{P}$. The idea of the Kusuoka's representation theorem is to write $\rho(X)$ as a supremum over distributions of probability densities with respect to $\mathbb{P}$.

[2] $\mathcal{L}^\infty$ is the vector space of bounded $\mathcal{F}$-measurable functions on Ω with the supremum norm

Proposition 4.4. *Let ρ be a law-invariant convex risk measure on $\mathcal{L}^\infty(\Omega, \mathcal{F}, \mathbb{P})$ that is continuous from above. Then its minimal penalty function is given by*

$$\gamma_\rho(Q) = \sup_{X \in \mathcal{A}_\rho} \int_0^1 VaR_\lambda(X) q_{dQ/d\mathbb{P}}(1-\lambda) d\lambda$$

where $q_{dQ/d\mathbb{P}}$ is any quantile function of the distribution of $dQ/d\mathbb{P}$.

Proof. Using the law-invariance of ρ and the fact that it is equivalent that if $(\Omega, \mathcal{F}, \mathbb{P})$ is atomless, i.e., for every $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$, there exists $A' \in \mathcal{F}$ with $A' \subset A$ and $0 < \mathbb{P}(A') < \mathbb{P}(A)$, there exists a standard uniform random variable U on the probability space, and for every real-valued random variable X there exists a standard uniform random variable U such that $X = q_X(\mathrm{U})$.[3]

$$\begin{aligned}\gamma_\rho(Q) &= \sup_{X \in \mathcal{A}_\rho} E_Q[-X] \\ &= \sup_{X \in \mathcal{A}_\rho} \sup_{\widetilde{X} \sim X} E_Q[-\widetilde{X}] \\ &= \sup_{X \in \mathcal{A}_\rho} \sup_{\widetilde{X} \sim X} E_Q[-\widetilde{X}\frac{dQ}{d\mathbb{P}}] \\ &= \sup_{X \in \mathcal{A}_\rho} \int_0^1 q_{-X}(\lambda) q_{dQ/d\mathbb{P}}(\lambda) d\lambda.\end{aligned}$$

Moreover, $q_{-X}(\lambda) = -q_X(1-\lambda) = VaR_{1-\lambda}(X)$. □

Corollary 4.1. *$\gamma_\rho(Q)$ depends only on the distribution of $dQ/d\mathbb{P}$ under $\mathbb{P}$.*

Theorem 4.1 (Kusuoka's representation theorem). *Let ρ be a law-invariant convex risk measure on $\mathcal{L}^\infty(\Omega, \mathcal{F}, \mathbb{P})$ that is continuous from above. Then*

$$\rho(X) = \sup_{\mu \in \mathcal{M}_1((0,1])} \left(\int_{(0,1]} AVaR_\lambda(X)\mu(d\lambda) - \gamma_\rho(\mu) \right)$$

where $\gamma_\rho(\mu) = \sup_{X \in \mathcal{A}_\rho} \int_{(0,1]} AVaR_\lambda(X\mu(d\lambda))$.

[3] If X and Y are two random variables on an atomless probability space, and assume that one of them is integrable and the other is bounded. Let U be a standard uniform random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ s.t. $X = q_X(\mathrm{U})$ $\mathbb{P}$-a.s. Then the maximum is attained by $\widetilde{Y} = q_Y(\mathrm{U})$, and consequently

$$\sup_{\widetilde{Y} \sim Y} E[X\widetilde{Y}] = E[q_X(\mathrm{U}) q_Y(\mathrm{U})] = \int_0^1 q_X(t) q_Y(t) dt.$$

Proof. The right-hand side of the equation defines a law-invariant convex risk measure that is continuous from above. Conversely, let ρ be a law-invariant and continuous from above. We will show that for $Q \in \mathcal{M}_1(\mathbb{P})$ there exists a measure $\mu \in \mathcal{M}_1((0,1])$ s.t.

$$\int_0^1 q_{-X}(t) q_{dQ/d\mathbb{P}}(t) dt = \int_{(0,1]} AVaR_s(X)\mu(ds).$$

Then, since $q_{-X}(t) = VaR_{1-t}(X)$ and $q_{dQ/d\mathbb{P}}(t) = q^+_{dQ/d\mathbb{P}}(t)$ for $t \in (0,1)$,

$$\int_0^1 q_{-X}(t) q_{dQ/d\mathbb{P}}(t) dt = \int_0^1 VaR_t(X) q^+_{dQ/d\mathbb{P}}(1-t) dt.$$

Since $q^+_{dQ/d\mathbb{P}}$ is increasing and right-continuous, we can write $q^+_{dQ/d\mathbb{P}}(t) = \nu\big((1-t,1]\big)$ for some positive locally finite measure ν on $(0,1]$. Moreover, the measure μ given by $\mu(dt) = t\nu(dt)$ is a probability measure on $(0,1]$:

$$\int_{(0,1]} t\nu(dt) = \int_0^1 \nu((s,1]))ds = \int_0^1 q^+_{dQ/d\mathbb{P}}(s) ds = E[dQ/d\mathbb{P}] = 1$$

Thus,

$$\begin{aligned}\int_0^1 q_{-X}(t) q_{dQ/d\mathbb{P}}(t) dt &= \int_0^1 VaR_t(X)\big(\int_{(t,1]} \frac{1}{s}\mu(ds)\big)dt \\ &= \int_{(0,1]} \frac{1}{s}\int_0^s VaR_t(X)dt\mu(ds) \\ &= \int_{(0,1]} AVaR_s(X)\mu(ds).\end{aligned}$$

Conversely, for any probability measure μ on $(0,1]$, the function q defined by $q(t) := \int_{(1-t,1]} s^{-1}\mu(ds)$ can be viewed as the quantile function of the density $dQ/d\mathbb{P} = q(U)$ of a measure $Q \in \mathcal{M}_1(\mathbb{P})$, where U has a uniform distribution on $(0,1)$. Altogether, we obtain a one-to-one correspondence between laws of densities $dQ/d\mathbb{P}$ and probability measures μ on $(0,1]$. □

4.4 Conic Finance

In this last section, we will introduce the theory of Conic Finance. The markets for the relatively liquid assets of an economy are modelled in classical finance as a counterparty for market participants, and they are seen as accepting any amount and direction of financially traded asset at the going market price. Particular forms of liquid assets describe the cash flow to be accessed for which there is just one price. However, when modelling markets more broadly, we allow prices to vary

with the trade direction. Hence, there are two prices, one for buying from the market (***ask price***), and another one for selling to the market (***bid price***) (Madan and Cherny [22]). In a modern financial economy, all risks cannot be eliminated. Perfect hedging is not possible and some risk exposures must be tolerated. Hence the set of acceptable risks must be defined as a financial primitive of the financial economy. The theory is founded on the basis of the concepts of acceptability of stochastic cash-flows and distorted expectations.
Following Madan and Schoutens [21], suppose that a non-negative random variable X, a zero cost stochastic cash-flow at some particular time (T), is identified on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Even though the assumption of a zero cost investment is not entirely realistic, it does not affect the generality of the theory, as the premium can be borrowed at a risk-free rate and paid back at the final pay-off date of the cash-flow. The basic examples we will work with are $Y = X - \exp(rT)b$ or $Z = \exp(rT)a - X$, where r is the interest rate at date T. Finding $X - \exp(rT)b$ acceptable is then seen as the market being willing to buy the risk X for the initial price b (the other case refers to selling at the initial price a). Consider now the value of a financial derivative in the traditional one-price setting, which is equal to the discounted risk-neutral expectation of the pay-off X:

$$V(X) = \exp(-rT)E_Q[X]$$

Where Q is the risk-neutral measure [4]. Now, the value of the accepted price Y is

$$V(Y) = \exp(-rT)E_Q[X - \exp(rT)b] = V(X) - b \geq 0, \text{ for } b \leq V(X)$$

We can say that we are now actually selling X to the market for a price b, below the risk-neutral market price $V(X)$. We can build an analogous reasoning for the price Z, for which we would be buying X for a price above the risk-neutral value. First we can consider the set of zero-cost cash-flows, defined as

$$\mathcal{A}^* = \{W \mid V(W) \geq 0\},$$

as a potential set of admissible risks. It is the largest possible convex cone containing the non-negative random variables: these are always acceptable as they are arbitrage. However, the set of acceptable zero-cost cash-flows, $\mathcal{A} \subset \mathcal{A}^*$, of a two-price economy will more precisely be a proper convex set containing the non-negative random variables. It is defined as

$$W \in \mathcal{A} \Longleftrightarrow \exp(-rT)E_Q[W] \geq 0 \text{ for all } Q \in \mathcal{M}$$

[4] A risk-neutral measure is a probability measure such that each share price is exactly equal to the discounted expectation of the share price under this measure.

Both $Y = X - \exp(rT)b$ and $Z = \exp(rT)a - X$ are in $\mathcal{A}$. That is to say,

$$\exp(-rT)E_Q[X - \exp(rT)b] = \exp(-rT)E_Q[X] - b \geq 0$$
$$\exp(-rT)E_Q[\exp(rT)a - X] = a - \exp(-rT)E_Q[X] \geq 0$$

Then, the best bid and ask prices for X provided by the market are given by

$$bid(X) = \exp(-rT) \inf_{Q \in \mathcal{M}} E_Q[X]$$
$$ask(X) = \exp(-rT) \sup_{Q \in \mathcal{M}} E_Q[X]$$

Every market is defined by a convex cone of zero cost cash-flows acceptable to the market, and this cone has associated with it a convex set of probability measures $Q \in \mathcal{M}$ with acceptability equivalently defined as positive expectation under each Q. Therefore, the financial markets for the law of two prices are referred to as ***conic***.

Definition 4.8. *A concave distortion function is a concave function $\psi(x)$ from the unit interval to itself*

$$\psi : [0,1] \to [0,1]$$
$$x \mapsto \psi(x)$$

If we restrict acceptability to be defined completely by the probability distributions of the associated risk, then the calculation of bid and ask prices can be made quite tractable using such concave distortion functions. We will assume that these risks satisfy the comonotonicity[5] and additivity[6] properties. Following the results in Kusuoka [20], under these hypotheses, bid and ask prices must be expectations under a concave distortion. That is, the bid price is given by

$$bid(X) = \exp(-rT) \int_{-\infty}^{+\infty} x d\psi(F_X(x))$$

and the ask price is

$$ask(X) = -\exp(-rT) \int_{-\infty}^{+\infty} x d\psi(F_{-X}(x))$$

[5] Two risks X and Y, are said to be comonotone if they are actually completely driven by one single risk factor. That is, there exists a random variable, Z, on the unit interval s.t. $X = F_X^{-1}(Z)$ and $Y = F_Y^{-1}(Z)$, where F_X^{-1} and F_Y^{-1} denote the inverse of the distribution of X and Y, respectively.

[6] We assumed coherent risk measures to be sub-additive (see axiom 4.6). For two random variables X and Y, $bid(X+Y) \geq bid(X) + bid(Y)$ and $ask(X+Y)) \leq ask(X) + ask(Y)$. These relations can be seen as the effect of diversification, but when these variables are comonotone, we could say there is no scope for diversification. Hence, for such pair of random variables, we require in fact that $bid(X+Y) = bid(X) + bid(Y)$ (resp. *ask*).

A random cash-flow will have a high acceptability level if its distribution function withstands high levels of stress. That is, if its expectation is still positive after distorting its distribution function. The level of acceptability is then proportional to the level of stress.

Definition 4.9. *An index of acceptability is a map α defined as follows*

$$\begin{aligned} \alpha : \mathcal{X} &\to [0, +\infty] \\ X &\mapsto \alpha(X) \end{aligned}$$

The number $\alpha(X)$ is the level of acceptability of X.

The map α is an acceptability index if, and only if, there exists an increasing one-parameter family of coherent risk measures $\{\rho_\lambda, \lambda \geq 0\}$ (see 4.1) , with the property that $\alpha(Y)$ is the largest level λ such that the cash-flow Y is acceptable to the level λ:

$$\alpha(Y) = \sup\{\lambda \geq 0 \mid Y \in \mathcal{A}_\lambda\}$$

where

$$Y \in \mathcal{A}_\lambda \Longleftrightarrow \exp(-rT)E_Q[Y] \geq 0, \text{ for all } Q \in \mathcal{M}_\lambda$$

with $\mathcal{M}_\lambda$ is the convex set of probability measures associated with the coherent risk measure ρ_λ. If we have a given bid or ask price, and the distribution function of the related cash flow, we could calculate the particular λ that one needs to use in the given distortion to obtain the given price.
To end with this section, we will give some examples of distortion functions used as an operational tool to calculate bid and ask prices.

Example 4.6 (Examples of distortion functions)**.** *The following functions are some of the most common distortion functions used in finance:*

i) MINVAR: $\psi_\lambda^{MINVAR}(x) = 1 - (1-x)^{1+\lambda}$.

ii) MAXVAR: $\psi_\lambda^{MAXVAR}(x) = x^{\frac{1}{1+\lambda}}$.

iii) MAXMINVAR: $\psi_\lambda^{MAXMINVAR}(x) = (1 - (1-x)^{1+\lambda})^{\frac{1}{1+\lambda}}$

iv) MINMAXVAR: $\psi_\lambda^{MINMAXVAR}(x) = 1 - (1 - x^{\frac{1}{1+\lambda}})^{1+\lambda}$

v) Wang Transform: $\psi_\lambda^{WANG}(x) = N(N^{-1}(x) + \lambda)$, *which is defined using the cumulative distribution function of the standard normal N(x) and its inverse function.*

Conclusions

Throughout this work, we have established how preferences and its quantification through utility functions allow us to set up a decision-making model that axiomatises rational choice and risk bearing. Together with logical analysis and empirical research on behaviour and risk attitudes, we have been able to build some mathematical structures that are remarkably useful in finance, as they lead to an increased efficiency in investments and markets. I would like to highlight the third chapter. In that chapter, we have presented many drawbacks of the EUT, but we have been able to adapt and extend the model to fit subjective distortions, which in the end has led us to numerical representations for robust preferences. And by working with the latter, we have given a proper approach to monetary risk measures and Conic Finance, which gives us the possibility to understand how prices are formed in financial markets.

Before studying Mathematics, I studied Economics at the UB. There, what captivated me the most were Microeconomics and Game Theory. I was very interested in the behaviour of individuals in making decisions that involve risk, and how these individuals interacted with each other. This work has given me the opportunity to get a deeper understanding in this field, and has also been very useful to retake many subjects that I presumed forgotten. Overall, I have genuinely enjoyed the research, as I find fascinating the fact that we are able to model human behaviour towards choice with maths.

Bibliography

[1] Maurice Allais. Allais paradox. In *Utility and probability*, pages 3–9. Springer, 1990.

[2] Maurice Allais and GM Hagen. *Expected utility hypotheses and the Allais paradox: Contemporary discussions of the decisions under uncertainty with Allais' rejoinder*, volume 21. Springer Science & Business Media, 2013.

[3] Kenneth J Arrow. *Social choice and individual values*, volume 12. Yale university press, 2012.

[4] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.

[5] Daniel Bernoulli. Exposition of a new theory on the measurement of risk. In *The Kelly capital growth investment criterion: Theory and practice*, pages 11–24. World Scientific, 2011.

[6] So Yeon Chun, Alexander Shapiro, and Stan Uryasev. Conditional value-at-risk and average value-at-risk: Estimation and asymptotics. *Operations Research*, 60(4):739–756, 2012.

[7] Gerard Debreu. Continuity properties of paretian utility. *International Economic Review*, 5(3):285–293, 1964.

[8] John Eatwell, Murray Milgate, and Peter Newman. *Utility and probability*. Springer, 1990.

[9] Louis Eeckhoudt, Christian Gollier, and Harris Schlesinger. *Economic and financial decisions under risk*. Princeton University Press, 2011.

[10] Allan M Feldman and Roberto Serrano. *Welfare economics and social choice theory*. Springer Science & Business Media, 2006.

[11] Peter C Fishburn. The irrationality of transitivity in social choice. *Behavioral Science*, 15(2):119–123, 1970.

[12] Peter C Fishburn. Nontransitive preferences in decision theory. *Journal of risk and uncertainty*, 4(2):113–134, 1991.

[13] Hans Follmer and M Scheizer. Encyclopedia of quantitative finance, 2010.

[14] Hans Föllmer and Alexander Schied. *Stochastic finance: an introduction in discrete time*. Walter de Gruyter, 2011.

[15] Andreas H Hamel et al. Monetary measures of risk. Technical report, 2018.

[16] John R Hauser and Glen L Urban. Assessment of attribute importances and consumer utility functions: Von neumann-morgenstern theory applied to consumer behavior. *Journal of Consumer Research*, 5(4):251–262, 1979.

[17] Israel N Herstein and John Milnor. An axiomatic approach to measurable utility. *Econometrica, Journal of the Econometric Society*, pages 291–297, 1953.

[18] Daniel Kahneman, Jack L Knetsch, and Richard H Thaler. Experimental tests of the endowment effect and the coase theorem. *Journal of political Economy*, 98(6):1325–1348, 1990.

[19] Botond Kőszegi and Matthew Rabin. A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165, 2006.

[20] Shigeo Kusuoka. On law invariant coherent risk measures. In *Advances in mathematical economics*, pages 83–95. Springer, 2001.

[21] Dilip Madan and Wim Schoutens. *Applied conic finance*. Cambridge University Press, 2016.

[22] Dilip B Madan and Alexander Cherny. Markets as a counterparty: an introduction to conic finance. *International Journal of Theoretical and Applied Finance*, 13(08):1149–1177, 2010.

[23] Andreu Mas-Colell, Michael Dennis Whinston, Jerry R Green, et al. *Microeconomic theory*, volume 1. Oxford university press New York, 1995.

[24] Oskar Morgenstern and John Von Neumann. *Theory of games and economic behavior*. Princeton university press, 1953.

[25] Lars Tyge Nielsen. Monotone risk aversion. *Economic Theory*, 25(1):203–215, 2005.

[26] John W Pratt. Risk aversion in the small and in the large. In *Uncertainty in Economics*, pages 59–79. Elsevier, 1978.

[27] Matihew Rabin. Risk aversion and expected-utility theory: A calibration theorem. In *Handbook of the fundamentals of financial decision making: Part I*, pages 241–252. World Scientific, 2013.

[28] Stephen A Ross. Some stronger measures of risk aversion in the small and the large with applications. *Econometrica: Journal of the Econometric Society*, pages 621–638, 1981.

[29] Ariel Rubinstein et al. Lecture notes in microeconomic theory. *Princeton, NJ: Princeton University*, 2006.

[30] Paul A Samuelson. A note on the pure theory of consumer's behaviour. *Economica*, 5(17):61–71, 1938.

[31] Leonard J Savage. *The foundations of statistics*. Courier Corporation, 1972.

[32] Uzi Segal. The ellsberg paradox and risk aversion: An anticipated utility approach. *International Economic Review*, pages 175–202, 1987.

[33] Kotaro Suzumura. Rational choice and revealed preference. *The Review of Economic Studies*, 43(1):149–158, 1976.

[34] Richard Thaler. Toward a positive theory of consumer choice. *Journal of economic behavior & organization*, 1(1):39–60, 1980.

[35] Amos Tversky and Daniel Kahneman. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.

[36] Amos Tversky and Daniel Kahneman. The framing of decisions and the psychology of choice. *science*, 211(4481):453–458, 1981.

[37] Susheng Wang et al. *Microeconomic theory*. Springer, 2018.