Advanced Microeconomics 1

Juuso Välimäki

September 2024

These lecture notes are written for a research M.Sc. course in microeconomic theory covering the theory of individual choice. They are meant to complement the course textbook *Microeconomic Theory* by Mas-Colell, Whinston and Green, and the material presented in the lectures. Special thanks to Mikael Mäkimattila for useful comments on a previous version.

Introduction

The four parts in the research M.Sc. sequence in microeconomic theory at Helsinki GSE cover: Decision Theory (Part I), Welfare Economics and Competitive Markets (Part II), Game Theory (Part III) and Economics of Information (Part IV). At a very general level, the aim of Part I is to introduce formal models of individual decision making. This means that the course has two goals: it gives the students a first look at decision theoretic models and one goal is to present a variety of substantive economic models. The second goal is to introduce students to formal arguments. This explains why the notes may seem overly pedantic at times.

Key concepts in this course are choice rules and preference relations on various domains, utility representations of preference relations, and implications of utility maximization in different contexts. Part II analyzes collective choice problems (with multiple economic agents) and market models of resource allocation. Part III extends the analysis to cover strategic aspects, and Part IV concentrates on models of imperfect and incomplete information.

These lecture notes are organized as follows:

1. Rational Choice
   (a) Choice Rules: Coherent Choice and Independence of Irrelevant Alternatives
   (b) Preferences: Revealed Preference from Coherent Choice Rules and Choice Rules from Rational Reference Relations
   (c) Utility Representations: Real-Valued Functions Representing Preferences
2. Revealed Preference
   (a) Classical Consumer Choice via Revealed Preference
   (b) Firm's Problem via Revealed Profit
3. Maximizing a Numerical Objective Function
   (a) Continuous Utility Representation and Choice from Budget Sets
   (b) KKT Conditions for Utility and Profit Maximization
   (c) Comparative Statics of Optimal Choice
   (d) Value Functions and Envelope Theorem
4. Duality in Consumer and Firm Problems
   (a) Expenditure Minimization, Slutsky Equation, and Integrability
   (b) Cost Function, Shephard's Lemma, Hotelling's Lemma
5. Choice Under Uncertainty
   (a) Domain of Choice: Lotteries

   (b) Expected Utility Theorem
   (c) Monetary Payoffs: Risk Aversion and Comparison of Risks
   (d) Risk and Time

6. Advanced Topics:
   (a) Probability Weighting, Rank-Dependent Utility
   (b) Behavioral Theories: Time-Inconsistent Preferences, Reference-Dependent Utility
   (c) Menus of Lotteries: Preference for Flexibility, Temptation
   (d) Stochastic Choice

I have included in some sections material that goes beyond the scope of this course in order to help the interested student to get a broader picture of the current state of choice and decision theory. Obviously this course cannot teach the advanced topics properly, and the idea is to give a flavor of the questions analyzed. I also give some suggestions for further reading. These advanced topics are ideal material for a course essay. I have tried to mark these sections in such a way that it is easy to tell them apart from the required material .

1 Rational Choice

A decision maker chooses from a fixed sets of alternatives. For example, the choice could be over different locations for a weekend holiday in October. One set of alternatives could be: $\{$Barcelona, Rome, London, Paris$\}$. If we present this set of options to the agent, we will observe a single choice from the set.

Different decision makers will choose differently from this set, but we would not conclude that some of the agents are inconsistent or irrational. They are just different. In order to judge the consistency of a decision maker, we must observe her choices in a range of different decision problems.

The objective of this first part of the course is to give an introduction to various ways of thinking about rational choice. The first approach, *choice-based theory* relies only on observed choices made from different sets of alternatives. The second approach, *preference-based theory* assumes that choice behavior arises from pairwise comparisons between alternatives. The third approach, *utility-based theory* assumes that the decision maker has a numerical criterion that she can use in comparing possible choices.

Most economists would agree that the choice-based approach is the most primitive one. In that approach, the analyst does not impose any structure to unobservable motives as in the preference-based approach and the utility-based approach. The purpose of this first part of the course is to convince you that under reasonable assumptions on consistency or coherence of choice behavior, these three approaches result in remarkably similar analysis.

In undergraduate studies (in intermediate microeconomics and mathematics for economists), consumer theory is sometimes presented as follows: specify a utility function for the decision maker and use constrained optimization to derive the resulting demand behavior. If you think about this, it is a bit funny way to proceed. Are we supposed to test this with all possible utility functions?

The approach in this course asks: What can we say about coherent or rational choice behavior in general. What are the empirical patterns that falsify our model of choice? From this point of view, you can ask if demand behavior of the form where a constant income share is spent on each good is compatible with our notion of coherence. We will see that one way of answering this question is by showing that such demand behavior satisfies the sufficient conditions for rationalizability by preference maximization. In fact (as you probably know from Intermediate Microeconomics), these demands result from the maximization of Cobb-Douglas -preferences. A key goal in this course is the development of criteria based on observable behavior that allow us to determine when choices are made in a coherent manner.

1.1 Coherent Choice Functions

1.1.1 Single-Valued Choice Functions

We start with a domain of alternatives that could conceivably be available for the economic agent. We denote this set by X. A choice problem is any nonempty subset $A \subset X$. A *choice function* is a function $c : 2^X \setminus \emptyset \to A$,[1] and we denote the choice from set A by $c(A) \in A$.[2]

How should we think of consistent choice? Suppose we know that $c(A) = a$, i.e. that option a was chosen from a set of alternatives A. Suppose that the decision maker is presented with a subset of the original options $B \subset A$. Since all the options contained in B were already available when choosing from A, and since a was chosen from that larger set, we might expect that if $a \in B$, then $c(B) = a$. This is formalized in the

[1] The notation in these notes is as follows: $f : X \to Y$ denotes a function that has the set X as its domain, and Y as its codomain. We denote the set of all subsets of X by 2^X, and therefore $2^X \setminus \emptyset$ is the set of all non-empty subsets of X

[2] Notice that we are restricting choice rules to be single-valued. We consider in the next subsubsection *choice correspondences* where we allow for $c(A) \subset A$, with $c(A) \neq \emptyset$ for all non-empty $A \subset X$.

following definition attributed variously to Arrow, Chernoff and Radner & Marschak (and also known as Sen's α):

Definition 1.1 (Independence of Irrelevant Alternatives, IIA)**.** If $c(A) = a$, and $a \in B \subset A$, then $c(B) = a$.

This condition simply requires that by deleting irrelevant alternatives, i.e. elements not chosen, the choice amongst remaining alternatives will not change. In the context of our original example, if Barcelona is chosen from the set {Barcelona, Rome, London, Paris}, then Barcelona should be the choice from the set {Barcelona, London, Paris} as well. This sounds like an innocent enough requirement, and IIA is the main assumption for rational choice functions that we impose. A choice function satisfying IIA is often called coherent, and the first part of this course is devoted to the study of coherence choice functions. [3].

1.1.2 Çhoice Correspondences

Suppose next that $c(A)$ can be any non-empty subset of A. We call such $c(A)$ *choice correspondences*. The interpretation of this case is that the decision maker could have chosen any of the elements in $c(A)$. We will see that when interpreting choice through the lens of a decision maker's preferences, it is natural to allow such set-valued choices.

In this setting, Sen's α takes a slightly different form to account for the possibility of multi-valued choice sets:

Definition 1.2. A Choice correspondence $c(A)$ satisfies Sen's α if $a \in B \subset A$ and $a \in c(A)$ implies $a \in c(B)$.

[3]There are numerous famous examples where IIA fails in experiments. Notable examples include: framing effects (different choices depending on different descriptions of the same alternatives lead to different choices), and decoy effects(the whole set of alternatives affects comparisons between alternatives $a \in A$ and $b \in A$). Osborne and Rubinstein Subsection 2.6 discuss such instances

For choice correspondences, another assumption is made to describe coherent choices. It is called Sen's β:

Definition 1.3 (Sen's β)**.** If $a, b \in c(B)$, $B \subset A$, and $a \in c(A)$, then $b \in c(A)$.

Sen's α is a restriction on choice behavior when contracting the set of alternatives. Sen's β puts conditions on an expanding choice set. I note here that there is an equivalent way of stating Sen's α and β (and this will be useful later).

Definition 1.4 (WARP)**.** A choice correspondence satisfies *weak axiom of revealed preference* if $a, b \in A \cap B$ and $a \in c(A)$ and $b \in c(B)$ implies $a \in c(B)$.

Exercise 1.5. Prove that c satisfies WARP if and only if it satisfies Sen's α and Sen's β.

Let me end this subsection with a comment on connections. Stochastic choice theory could be thought as a special case of choice correspondences. This theory analyzes probabilistic choice from a set of alternatives, in particular the probability of choosing alternative $a \in A$ denoted by $\rho(a, A)$ as a function of the choice set A. This literature has connections to empirical models of discrete choice (as in industrial organization or structural labor economics). The book by Tomasz Strzalecki on the syllabus is an excellent source on this topic, and I also include a brief introduction to stochastic choice in the last section of these notes.

1.2 Rational Preferences

An alternative way to think about choice is via pairwise comparisons. For this, we assume that our decision maker can compare any $x, y \in X$. By this we mean that there is a *binary relation* $\succeq$ on X. The interpretation of $\succeq$ is as follows: If $x \succeq y$, then $x \in X$ is considered at least as good as $y \in X$

Definition 1.6 (Properties of Binary Relations)**.** The binary relation $\succeq$ is called

1. Complete if $\forall\, x, y \in X,\ x \succeq y$ or $y \succeq x$ or both.

2. Transitive if $\forall\, x, y, z \in X,\ x \succeq y$ and $y \succeq z \implies x \succeq z$.

3. Reflexive if $\forall\, x \in X,\ x \succeq x$.

4. Antisymmetric if $\forall\, x, y \in X,\ x \succeq y$ and $y \succeq x \implies x = y$.

Exercise 1.7. Determine which of the above properties hold for $\geq$ on $\mathbb{R}$ and for $\geq$ on $\mathbb{R}^n$. Is any of the Properties implied by other properties on the list?

A binary complete and transitive binary relation is called a *weak order*. In microeconomics, we identify weak orders with *rational preferences*.

Definition 1.8. A binary relation $\succeq$ is called a *rational preference relation* if it is complete and transitive.

From a rational preference relation $\succeq$, we can derive two other binary relations. We say that x is strictly preferred to y if $x \succeq y$ and $\neg(y \succeq x)$ and write $x \succ y$. We say that x is indifferent to y if $x \succeq y$ and $y \succeq x$ and write $x \sim y$. I list below a few standard exercises that you may want to try in order to familiarize yourself with typical properties of binary relations.

Exercise 1.9. Show that if $\succeq$ is a rational preference relation, then $\succ$ (derived from $\succeq$ as above) is:

1. Asymmetric: If $x \succ y$, then $\neg(y \succ x)$.

2. Negatively transitive: If $\neg(x \succ y)$ and $\neg(y \succ z)$, then $\neg(x \succ z)$.

Exercise 1.10. Show that if $\succeq$ is a rational preference relation, then $\sim$ (derived from $\succeq$ as above) is an equivalence relation, i.e. it is:

1. Transitive.

2. Reflexive.

3. Symmetric: $\forall\, x, y \in X$, if $x \sim y$, then $y \sim x$.

Exercise 1.11. Show conversely that if $\succ$ is asymmetric and negatively transitive, then $\succeq$ derived from $\succ$ by:

$$x \succeq y \iff \neg(y \succ x),$$

is a rational preference relation.

If transitivity is violated, you can get cyclic preferences of the form $a \succ b \succ c \succ a$. Which is the most preferred alternative in this case? Cyclical preferences of this type could be exploited in a *money pump* as follows. Suppose the decision maker has possession of alternative a. Offer her the chance to switch to c for a small monetary amount. Then to b for a small monetary amount, then back to a again for a small monetary amount. At this point, she is back to the starting position, only poorer by three small amounts. Continue the process until the decision maker has no money left.

To see how cyclical preferences might arise, consider the following example.

Example 1.12. Suppose that you compare prospective apartments that you might rent based on three qualities: size, price, distance to Economicum. It is easy to rank the choices separately for each quality: bigger is better, cheaper is better, closer to Eco is better. Say each alternative is then a triple $a = (a_s, a_p, a_d)$ and similar for b. When you aggregate the preference over qualities into a preference over the alternatives, you say $a \succ b$ if and only if a is better than b on at least two qualities. This procedure can easily lead to violations of transitivity.

Exercise 1.13. A non-standard die is one where the numbers on each face are between 1 and 6 and sum up to 21. Let A denote the set of all such dice. Say that die a wins over die b in a fair roll of the dice if the number on the upward pointing face of a is larger than on the face of b. Define preference by $a \succ b$ if and only if $\Pr\{a \text{ wins}\} > \Pr\{b \text{ wins}\}$. Show by an example that this preference relation is not transitive. Is there any a such that $\neg(b \succ a)$ for all $b \in A$.

1.3 Connecting Choice and Preferences

In this subsection, I connect observable choices to a psychological motivation, the preferences. Let's do this first in the simplest possible setting where we look at choice functions and strict rational preferences defined on a finite domain X.[4]

1.3.1 Preferences and Choice Functions on a Finite Domain

The idea is simple: you should choose alternative a only if there are no alternatives b available such that $b \succ a$. Similarly, if you choose a when b is available, it makes sense to say that you think a is at least as good as b.

Despite the apparent simplicity of this construction, it leads to one of the most fundamental ideas in microeconomic theory: we will never know the true motives of decision makers. We can make judgments on their welfare etc. only to the extent that we can learn from their choices. Towards this goal, we will see if we can build a rational preference relationship that rationalizes the choices arising from a consistent choice function.

Let's start with the choice function $c(A)$ defined on all non-empty subsets of X. If $c(\{a, b\}) = a$, then we say that a is *revealed preferred* to all b, and we denote this by $a \succeq_c b$. Our goal is to show that if c satisfies IIA, then $\succeq_c$ is a rational preference relation.

Proposition 1.14. If c satisfies IIA on domain X, then the revealed preference relation $\succeq_c$ is an antisymmetric rational preference relation.

[4] This is equivalent to adding the requirement that the rational preference relation is antisymmetric.

Proof. To show the completeness of $\succeq_c$, we only need to note that since c is defined on all nonempty $A \subset X$, we have $c(\{a,b\}) = a$ or $c(\{a,b\}) = b$, i.e. either $a \succeq_c b$ or $b \succeq_c a$.

To show transitivity of $\succeq_c$, we assume that $a \succeq_c b$ and $b \succeq_c d$ for some $a, b, d \in A$. We need to show that whenever c satisfies IIA, we have $a \succeq_c d$.

By definition of $\succeq_c$, we have $c(\{a,b\}) = a$, and $c(\{b,d\}) = b$. Consider $c(\{a,b,d\})$. If $c(\{a,b,d\}) = d$, then $c(\{b,d\}) = b$ violates IIA. If $c(\{a,b,d\}) = b$, then $c(\{a,b\}) = a$ violates IIA. Hence we conclude $c(\{a,b,d\}) = a$ and by IIA this implies that $c(\{a,d\}) = a$, i.e. that $a \succeq_c d$.

To show antisymmetry, if $a \succeq_c b$ and $b \succeq_c a$, then $a = c(\{a,b\}) = b$, i.e. $a = b$ since c is single-valued. □

The other direction is in fact quite simple. If we have a finite X, and if preferences are rational and antisymmetric, then we can find for all $A \subset X$ a unique element $a \in A$ such that $\forall\, b \neq a,\ b \in A,\ a \succeq b$. We call such an a the $\succeq$-maximal element in A and denote it by $c_{\succeq}(A)$.

Lemma 1.15. If X is finite, and $\succeq$ is a rational and antisymmetric preference relation, then $c_{\succeq}(A)$ exists.

Proof. We need to show that for all A, there is an $a' \in A$ such that $a' \succeq a$ for all $a \in A$.

Since A is finite, we can order it as $A = \{a_1, ..., a_n\}$ for some $n < \infty$. Set $a'_1 = a_1$.

To get a'_k for $k > 1$, compare a'_{k-1} and a_k. If $a'_{k-1} \succeq a_k$, set $a'_k = a'_{k-1}$. If $a_k \succeq a'_{k-1}$, set $a'_k = a_k$. By antisymmetry, only one of these cases is true. We claim that $a'_n \succeq a$ for all $a \in A$.

To see this, pick any $a_k \in A$. Since $a'_k \succeq a'_{k-1}$ for all $k > 1$ and $a'_k \succeq a_k$, we have that $a'_n \succeq a_k$ (by $n > k$ and by the transitivity of $\succeq$).

To see uniqueness, suppose that for some $a, a' \in A$, both $a \succeq b$ for all $b \in A$ and $a' \succeq b$ for all $b \in A$. Then antisymmetry implies that $a = a'$, i.e. the $\succeq$-maximal element is unique. □

Can you find an example of a set X (not finite) with a complete, transitive and antisymmetric binary relation on it such that no $\succeq$-maximal element exists? Apart from this existence problem, there is really no reason to insist on finite X. But note how we used the fact that the choice function was defined on all subsets of X.

Proposition 1.16. If $\succeq$ is a rational and antisymmetric preference relation, then $c_{\succeq}(A)$ is a choice function that satisfies IIA.

Proof. By the previous lemma, $c_{\succeq}(A)$ is non-empty and single-valued for all non-empty $A \subset X$. Hence we only need to show that it satisfies IIA. If $c_{\succeq}(A) \in B \subset A$, then $c_{\succeq}(A) \succeq b$ for all $b \in B$ and therefore $c_{\succeq}(A) = c_{\succeq}(B)$ and $c_{\succeq}(A)$ satisfies IIA. □

The following exercise completes the story. If we start with a choice function and derive its revealed preference relation and finally derive the choice function induced by the revealed preference relation, we get back the original choice function.

Exercise 1.17. Show that if $c(A)$ satisfies IIA, then $c_{\succeq_c}(A) = c(A)$.

I should mention that there is some debate amongst economists regarding the status of preference relations. Some view these as a convenient mathematical representations for coherent choice functions. Others believe that preferences are substantive, i.e. they describe true psychological deliberations that guide decision makers in choice situations. In these lectures, we do not need to take a stance on this, but e.g. the handbook chapter by Robson and Samuelson (2010) on the syllabus contains a good discussion around this topic.

1.3.2 Choice Correspondences and Preferences

The material in this section allows for a direct connection to classical consumer theory. The details are slightly more cumbersome, but the main idea is the same: define a revealed preference relation and then show that the choice induced by this revealed preference relation coincides with the original choice correspondence.

We will now drop the requirement of antisymmetry of preference relations to cope with multi-valued choice sets. When we drop antisymmetry, we allow for indifference between different alternatives. With this change, there is at least a possibility that the set of $\succeq$-maximal elements from A might coincide with the value of the choice correspondence at A.

In order to avoid problems with existence, we assume that the choice function is defined on all finite (non-empty) subsets of X.[5] It turns out that the weak axiom of revealed preference is exactly the condition needed to connect the two approaches for finite choice sets.

Definition 1.18 (WARP)**.** A choice correspondence satisfies *weak axiom of revealed preference* if $a, b \in A \cap B$ and $a \in c(A)$ and $b \in c(B)$ implies $a \in c(B)$.

Let's define the choice function induced by a rational preference relation.

Definition 1.19. Let $\succeq$ be a rational preference relation. The choice correspondence $c_{\succeq}(A)$ defined by $a \in c_{\succeq}(A) \iff (\forall\, a' \in A,\ a \succeq a')$ is called the choice correspondence induced by $\succeq$.

We have the following proposition:

Proposition 1.20. The following statements are equivalent:

1. The choice correspondence c satisfies Sen's α and Sen's β.
2. The choice correspondence c satisfies WARP.
3. There is a rational preference relation $\succeq$ such that $c_{\succeq}(A) = c(A)$ for all finite $A \in X$.

[5]If we restrict to finite X, we do not need to worry about this. It is important that we assume that c is defined on a sufficiently rich class of subsets of X.

Proof. You have already proved that $1. \implies 2.$ in Exercise 1.5.

To prove that $2. \implies 3.$, assume that $c(A)$ satisfies WARP and let $a \succeq_c b \iff a \in c(a,b)$. Completeness follows from non-emptyness of $c(A)$ for all A. Transitivity of $\succeq_c$ follows if for all $a,b,c \in A$, we have: $a \in c(\{a,b\})$, and $b \in c(\{b,c\})$ imply $a \in c(\{a,c\})$. Consider $c(\{a,b,c\})$. There are three cases:

i) If $a \in c(\{a,b,c\})$, then $a \in c(\{a,c\})$ by Sen's α.

ii) If $b \in c(\{a,b,c\})$, then $b \in c(\{a,b\})$ by Sen's α, and thus $a \in c(\{a,b,c\})$ by Sen's β, and we are back to i).

iii) If $c \in c(\{a,b,c\})$, then $c \in c(\{a,c\})$ by Sen's α, and thus $b \in c(\{a,b,c\})$ by Sen's β, and we are back to case ii).

Hence $\succeq_c$ is rational.

We have by definition: $a \in c(A) \iff a \succeq_c b$ for all$b \in A$. If we define $c_{\succeq_c}(A) = \{a \in A | \neg(b \succ_c a \text{ for some } b \in A)\}$, we have $a \in c(A) \iff a \in c_{\succeq_c}(A)$.

I leave it as an exercise to show that $3. \implies 1.$

□

1.3.3 Summary For Now

Do not get lost in the details. The main point thus far is: a reasonable assumption on the observable side, i.e. IIA on choice functions, is observationally indistinguishable from a reasonable assumption on the unobservable side, i.e. a rational preference relation. In a sense, these two approaches shed light on what consistent choice might mean and if you accept one side, you accept the other (and if you reject one side, you reject the other). I will reserve further comments on the on the assumptions made until the end of this section.

1.4 Utility Representations

Let's connect all of this next to the main workhorse from Intermediate Microeconomics: Utility Maximization. What could be the most straightforward way to connect choice to utility maximization? Maybe we can assign numerical values to the alternatives and then we would just pick from any A the alternative with the highest numerical value.

Definition 1.21. We say that a utility function $u : X \to \mathbb{R}$ *represents the rational preference* relation $\succeq$ on X if for all $x, x' \in X$ we have $u(x) \geq u(x') \iff x \succeq x'$.

In the case of rational preference relationships on a finite domain X, this is very easy. Define a utility $u : X \to \mathbb{R}$ by:

$$u(x) = \#\{y \in X | x \succeq y\}.$$

Exercise 1.22. Show that with this definition of u, $x \succeq x' \iff u(x) \geq u(x')$.

An important observation is that since the utility function is based on preferences, and preferences are ordinal in nature (there is no measuring stick for how much one prefers x over y), utility functions also contain only ordinal information. This simply says that the numerical values of the utility function are irrelevant, the only economically meaningful information relates to the relative ranking of alternatives. In fact, we have:

Proposition 1.23. If $u : X \to \mathbb{R}$ represents $\succeq$ and $v : \mathbb{R} \to \mathbb{R}$ is strictly increasing, then $v(u) : X \to \mathbb{R}$ also represents $\succeq$.

Proof. If v is a strictly increasing function, then $x \geq x' \iff v(x) \geq v(x')$. So if u represents $\succeq$, then $x \succeq x' \iff u(x) \geq u(x') \iff v(u(x)) \geq v(u(x'))$. □

This is why for vectors $(x, y) \in X$ the utility function given by $u(x, y) = e^{\sqrt{x+2y}}$ represents the same preferences as $v(x, y) = x + 2y$. You will have seen this with e.g. Cobb-Douglas preferences: $u(x, y) = x^\alpha y^{1-\alpha}$ represents the same preferences as $v(x, y) = \alpha \ln x + (1 - \alpha) \ln y$.

The following is for those interested in generalizing the mathematical structure to cover the types of applications that will be covered in this class, in particular classical consumer theory.[6]

If X is countably infinite, let $X = \{x_1, x_2, ...\}$ be an enumeration of X. Then let:

$$u(x) = \sum_{i: x \succeq x_i} \frac{1}{2^i}.$$

If X is uncountably infinite, but has an order dense countable subset $Y = \{y_1, y_2, ...\} \subset X$, (i.e. for all $x, x' \in X$ such that $x \succ x'$, there is an $y \in Y$ such that $x \succeq y \succeq x'$), then this is modified to:

$$u(x) = \sum_{i: x \succeq y_i} \frac{1}{2^i}.$$

All of the cases that we will encounter in this course have order dense subsets. You will find a counterexample of rational preferences not representable by a utility function by Googling 'Lexicographic Preferences'.

We will return to the representation of preferences by continuous utility functions later in the course, but with more assumptions on the preferences.

1.5 Assessing Rational Choice

Let me add a few words on interpreting the previous findings. We may ask: 'What is the significance of finding a preference (and hence a utility) rationalization to a particular choice behavior?' In particular: suppose that c satisfies IIA in the simplest case without indifferences. Do we conclude that the decision maker is a utility maximizer?

[6]This is not important for understanding the main economic content and not used in the rest of these notes.

To my taste, the best thought experiment for this question is the following. Suppose that the alternatives in A are presented to the decision maker in a fixed order $\{a_1, a_2, ...a_K\}$. We say that the decision maker is *satisficing* if she has in mind an outside option $a^* \in A$ such that she will accept the first alternative $a_k \succeq a^*$ and she will choose a_K if $a^* \succ a_k$ for all $k \leq K$.

Clearly such a process does not fit our traditional view of rational decision making. A satisficing decision maker is content with anything that is good enough and does not seek for the best possible alternative. Nevertheless the choice function for this procedure satisfies IIA and therefore has a rationalization in terms of a rational antisymmetric rational preference relation.

The key take-away from this example and others like it is that rationalizability by rational preference relations does not mean that the mental process behind the choices conforms to our idea of typical rational decision making by a careful comparison of the available alternatives. It just means that we cannot tell the such satisficing behavior apart from behavior induced by rational preferences. Subsection 2.2 in Osborne and Rubinstein discusses various decision procedures and their relation to rational choice.

In this course, the main platform for evaluating the development in the lectures is in the Problem Sets. In general, the best way of testing your understanding of the material in the lectures is by doing the Problems and Exercises in Osborne and Rubinstein, MWG and in Rubinstein. Some of the problems are challenging so do not despair if you have difficulties with them.

2 Revealed Preference

Up to this point, our discussion of choice and preferences has been completely abstract and independent of any economic context. In this section, I connect the ideas of the previous chapter to the two main topics of intermediate microeconomics: consumers' and firms' choices in competitive markets.

2.1 Classical Consumer Theory via Revealed Preference

2.1.1 WARP for Budget Sets

Classical consumer theory considers decision makers with a fixed income w and L different consumption goods that can be purchased in varying quantities at fixed per unit prices $\boldsymbol{p} = (p_1, ..., p_L) \in \mathbb{R}^L_{++}$ (i.e. $p_l > 0$ for all $l \in \{1, ..., L\}$). In this case, we can identify the set of alternatives as the budget set $B(\boldsymbol{p}, w)$:

$$B(\boldsymbol{p}, w) = \{\boldsymbol{x} \in \mathbb{R}^L | \boldsymbol{p} \cdot \boldsymbol{x} = \sum_{l=1}^{L} p_l x_l \leq w\}.$$

Notice that here we have infinite sets of alternatives, but we assume that the consumer has a choice function, i.e. a unique choice from each possible budget set. In this setting, we write weak axiom of revealed preference as:

Assumption 2.1 (WARP′ for Budget Sets)**.** If $\boldsymbol{x}, \boldsymbol{x}' \in B(\boldsymbol{p}, w) \cap B(\boldsymbol{p}', w')$ and $\boldsymbol{x} \in c(B(\boldsymbol{p}, w))$ and $\boldsymbol{x}' \in c(B(\boldsymbol{p}', w'))$, then $\boldsymbol{x} \in c(B(\boldsymbol{p}', w'))$, i.e. $\boldsymbol{x} = \boldsymbol{x}'$.

The equality of $\boldsymbol{x}$ and $\boldsymbol{x}'$ follows from the fact that choice is assumed to be single-valued. An alternative way of stating this is in terms of the budget outlays directly. If we write the choice from budget set $B(\boldsymbol{p}, w)$ as $\boldsymbol{x}(\boldsymbol{p}, w)$, we can rewrite the axiom as follows:

Assumption 2.2 (WARP for Budget Sets)**.** If $\boldsymbol{x}(\boldsymbol{p}, w) \neq \boldsymbol{x}(\boldsymbol{p}', w')$, and $\boldsymbol{p}' \cdot \boldsymbol{x}(\boldsymbol{p}, w) \leq w'$, then $\boldsymbol{p} \cdot \boldsymbol{x}(\boldsymbol{p}', w') > w$.

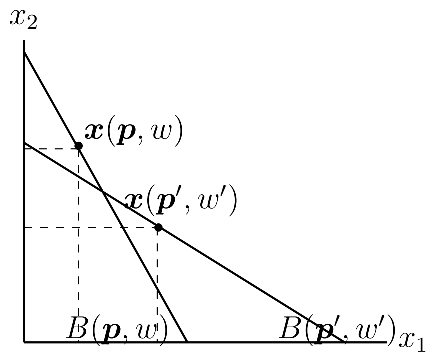


Figure 1: WARP i)

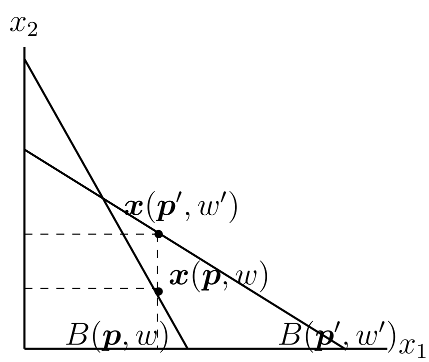


Figure 2: WARP ii)

Definition 2.3. If $\boldsymbol{p} \cdot \boldsymbol{y} \leq \boldsymbol{p} \cdot \boldsymbol{x}(\boldsymbol{p}, w)$, we say that $\boldsymbol{x}(\boldsymbol{p}, w)$ is *directly revealed preferred to* $\boldsymbol{y}$.

It is useful to draw illustrating pictures at this point for $\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2$.

Exercise 2.4. Determine which of the figures WARP i)-iii) are compatible with WARP.

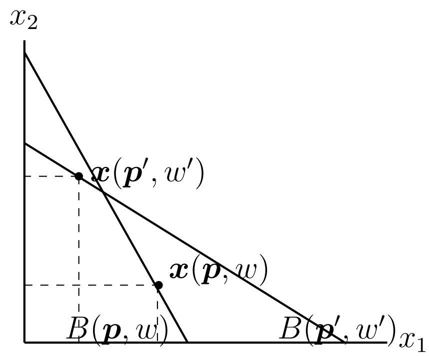


Figure 3: WARP iii)

2.1.2 WARP and the Compensated Law of Demand

WARP has surprisingly deep implications for consumer behavior. We start by observing the effect of price changes on choice that satisfies WARP. You should recall from intermediate microeconomics that for a utility maximizing consumer, the effect of any price change on consumption can be decomposed into a substitution effect and an income effect. In this deconposition, the substitution effect has a clear sign: if the price of a good rises, then substitution effect leads to a reduced demand. This same result holds for any demand satisfying WARP.

In order to get this result, we define an income compensation for a price change. The idea is simple: as prices change to say $\boldsymbol{p}'$, the outlay needed to purchase $\boldsymbol{x}(\boldsymbol{p}, w)$ changes. Let's adjust the decision maker's income to a new level w' so that $\boldsymbol{p}' \cdot \boldsymbol{x}(\boldsymbol{p}, w) = w'$.

Proposition 2.5. WARP implies the compensated law of demand, i.e. for $w' = \boldsymbol{p}' \cdot \boldsymbol{x}(\boldsymbol{p}, w)$, we have:

$$(\boldsymbol{p}' - \boldsymbol{p})(\boldsymbol{x}(\boldsymbol{p}', w') - \boldsymbol{x}(\boldsymbol{p}, w)) \leq 0.$$

In particular if only one price, say p_i changes, then $(p_i' - p_i)(x_i' - x_i) \leq 0$.

Proof. If $\boldsymbol{x}(\boldsymbol{p}', w') = \boldsymbol{x}(\boldsymbol{p}, w)$, the claim is trivially true so we assume that $\boldsymbol{x}(\boldsymbol{p}', w') \neq \boldsymbol{x}(\boldsymbol{p}, w)$ Since $\boldsymbol{x}(\boldsymbol{p}', w')$ is feasible, we have $\boldsymbol{p}' \cdot \boldsymbol{x}(\boldsymbol{p}', w') \leq w'$.

By the definition of compensation, $\boldsymbol{p}' \cdot \boldsymbol{x}(\boldsymbol{p}, w) = w'$ so that both $\boldsymbol{x}(\boldsymbol{p}, w)$ and $\boldsymbol{x}(\boldsymbol{p}', w')$ are feasible at the compensated budget set. Hence $\boldsymbol{x}(\boldsymbol{p}', w')$ is revealed preferred to $\boldsymbol{x}(\boldsymbol{p}, w)$. Since $\boldsymbol{x}(\boldsymbol{p}, w)$ satisfies WARP, it must be the case that $\boldsymbol{p} \cdot \boldsymbol{x}(\boldsymbol{p}', w') > \boldsymbol{p} \cdot \boldsymbol{x}(\boldsymbol{p}, w)$, and the claim follows. □

Let's make some further assumptions. Let's assume that $\boldsymbol{x}(\boldsymbol{p}, w)$ satisfies WARP and that the choice vector is a differentiable function of the prices and income. We have the compensated law of demand for infinitesimal compensated price changes $(d\boldsymbol{p}, dw)$, where $dw = \boldsymbol{x} \cdot d\boldsymbol{p}$:

$$d\boldsymbol{p} \cdot d\boldsymbol{x} \leq 0.$$

Write out the change in choice $\boldsymbol{x}$ from the price and the income change:

$$d\boldsymbol{x} = D_{\boldsymbol{p}}\boldsymbol{x}(\boldsymbol{p}, w)d\boldsymbol{p} + D_w\boldsymbol{x}(\boldsymbol{p}, w)dw.$$

$$= (D_{\boldsymbol{p}}\boldsymbol{x}(\boldsymbol{p}, w) + D_w\boldsymbol{x}(\boldsymbol{p}, w)\boldsymbol{x}(\boldsymbol{p}, w)^\top)d\boldsymbol{p}.$$

So we have:

$$d\boldsymbol{p} \cdot (D_{\boldsymbol{p}}\boldsymbol{x}(\boldsymbol{p}, w) + D_w\boldsymbol{x}(\boldsymbol{p}, w)\boldsymbol{x}(\boldsymbol{p}, w)^\top)d\boldsymbol{p} \leq 0.$$

The matrix $D_{\boldsymbol{p}}\boldsymbol{x}(\boldsymbol{p}, w) + D_w\boldsymbol{x}(\boldsymbol{p}, w)\boldsymbol{x}(\boldsymbol{p}, w)^\top$ is called the Slutsky matrix of the demand function $\boldsymbol{x}(\boldsymbol{p}, w)$. If you are not used to matrix notation with derivatives, it is a good exercise to write the above expression in terms of partial derivatives.

The result above tells us that the Slutsky matrix is negative semi-definite. We will see much later in this course that any demand function that is linearly homogenous in prices and income, uses the entire budget, and has a negative semi-definite and symmetric Slutsky matrix can be derived as a solution to a utility maximization problem. WARP does not imply symmetry of the Slutsky matrix, and this symmetry is not at all an obvious property that we would expect from well-behaved demands.

2.1.3 Loose Ends

I have derived some implications of WARP, but I am not aiming at a comprehensive treatment here. Let me point out some useful facts.

Result 1. Compensated law of demand (CLD) for all compensated price changes $(\Delta\boldsymbol{p}, \Delta w)$ implies WARP. The easiest way to get to this result is by considering the contrapositive: If WARP does not hold, then there is a compensated price change where CLD does not hold.

Exercise 2.6. Prove the above result.

We say that $\boldsymbol{x}$ is directly revealed preferred to $\boldsymbol{x}'$ if $\boldsymbol{x}$ is the choice at $(\boldsymbol{p}, w)$ and $\boldsymbol{p} \cdot \boldsymbol{x}' \leq w$, and we write $\boldsymbol{x} \succeq_{R^0} \boldsymbol{x}'$. We say that $\boldsymbol{x}$ is revealed preferred to $\boldsymbol{x}'$ and write $\boldsymbol{x} \succeq_R \boldsymbol{x}'$ if there is a sequence of choices $\boldsymbol{x}^1, ..., \boldsymbol{x}^n$ such that $\boldsymbol{x} \succeq_{R^0} \boldsymbol{x}^1 \succeq_{R^0} ... \succeq_{R^0} \boldsymbol{x}^n \succeq_{R^0} \boldsymbol{x}'$. (We say that $\succeq_R$ is the transitive closure of $\succeq_{R^0}$). We write $\boldsymbol{x} \succ_{R^0} \boldsymbol{x}'$ for $\boldsymbol{x}$ is directly revealed strictly preferred to $\boldsymbol{x}'$ if $\boldsymbol{x}$ is the choice at $(\boldsymbol{p}, w)$ and $\boldsymbol{p} \cdot \boldsymbol{x}' < w$. [7]

Definition 2.7. We say that a choice function $\boldsymbol{x}(\boldsymbol{p}, w)$ satisfies *strong axiom of revealed preference, SARP* if $\boldsymbol{x} \succeq_R \boldsymbol{x}'$ implies $\neg(\boldsymbol{x}' \succeq_{R^0} \boldsymbol{x})$. A choice correspondence $\boldsymbol{x}(\boldsymbol{p}, w)$ satisfies *generalized axiom of revealed preference, GARP* if $\boldsymbol{x} \succeq_R \boldsymbol{x}'$ implies $\neg(\boldsymbol{x}' \succ_{R^0} \boldsymbol{x})$.

Exercise 2.8. Show that the following demand observations satisfy WARP, but not SARP (or GARP)

$$\boldsymbol{p}^1 = (1, 1, 2),\ \boldsymbol{p}^2 = (2, 1, 1), \boldsymbol{p}^3 = (1, 2, 1 + \epsilon),$$

$$\boldsymbol{x}(\boldsymbol{p}^1, 1) = (1, 0, 0), \boldsymbol{x}(\boldsymbol{p}^2, 1) = (0, 1, 0), \boldsymbol{x}(\boldsymbol{p}^3, 1 + \epsilon) = (0, 0, 1).$$

Result 2 (Afriat's Theorem)**.** If the demand function $\boldsymbol{x}(\boldsymbol{p}, w)$ satisfies WARP and $L = 2$, then there is a utility function such that $\boldsymbol{x}(\boldsymbol{p}, w)$ is the solution

[7] The idea here is that if more is better, then $\boldsymbol{x}' + \epsilon\mathbf{1}$, with $\mathbf{1} = (1, ..., 1)$ is strictly preferred to $\boldsymbol{x}'$ and still feasible and thus $\boldsymbol{x}$ is directly revealed preferred to $\boldsymbol{x}' + \epsilon\mathbf{1}$.

to a utility maximization problem in $B(\boldsymbol{p}, w)$. If a choice function satisfies SARP and if a choice correspondence satisfies GARP, then they arise as solutions to utility maximization problems for some quasiconcave utility function.

Reny (2015) presents proofs and advanced discussions of these points. Varian (1982) shows how to use a finite set of observations $\{(\boldsymbol{p}^i, w^i, \boldsymbol{x}^i)\}_{i=1}^n$ to sketch the possible shapes of indifference curves for the decision maker.

2.2 Revealed Profitability

A firm uses inputs to produce outputs. In the theory of competitive firm behavior, we assume that the firm has no effect on market prices or at least takes input and output prices as exogenously given for its decisions.

2.2.1 Production Set

There are L goods in the economy. The goods can be used as inputs in the production process and some goods can be produced as output by the firm. The set of available alternatives for a competitive firm is called the *production set* and denoted by $Y \subset \mathbb{R}^L$. It is a list of technologically feasible vectors $y \in Y$. We adopt the convention that inputs in a production vector are represented by negative coordinates and outputs by positive coordinates. We will give more structure to the production set in Sections 3 and 4 when we talk about about profit maximizing competitive firms.

In classical firm theory, its objective is to maximize profit. For a competitive firm, profit is simply the revenue from output net of input costs. If both input and output prices are fixed and given by a price vector $p \in \mathbb{R}^L_+$, then the profit $\pi(y, p)$ from production vector $y \in Y$ is simply:

$$\pi(y) = p \cdot y = \sum_{l=1}^{L} p_l y_l.$$

Our perspective is now somewhat different to the consumer choice. We observe the firm's choices of inputs and outputs at various prices. Denote this collection of observed behavior by $\{(\boldsymbol{p}^i, \boldsymbol{y}^i)\}_{i=1}^n$. As outside observers, we do not know what the production set Y looks like, but we can estimate it based on the observations.

If $\boldsymbol{y}^i$ is chosen at price vector $\boldsymbol{p}^i$, we know that $Y \subset \{\boldsymbol{y} | \boldsymbol{p}^i \cdot \boldsymbol{y} \leq \boldsymbol{p}^i \cdot \boldsymbol{y}^i$. Can you see why this is the case? Obviously this can be done for all $i \in \{1, 2, ..., n\}$. This means that you have to take the intersection of the half-spaces below by the iso-profit hyperplanes that go through the observed production vectors.

Exercise 2.9. Assume that y_1 is an input and y_2 is an output. Find the region containing Y if the firm is a profit maximizer and has chosen the following production plans:

$$\{\boldsymbol{y}^i\} = \{(-2, 5), (-1, 3), (-5, 7)\} \text{ at prices } \{\boldsymbol{p}^i\} = \{(5, 2), (3, 2), (1, 1)\}.$$

2.2.2 The Law of Supply

We end this section with a comparative statics result. In our dataset $\{(\boldsymbol{p}^i, \boldsymbol{y}^i)\}_{i=1}^n$, the observed production vector generates maximal profit amongst the observed vectors, i.e. for all l, k:

$$\boldsymbol{p}^k \cdot \boldsymbol{y}^k \geq \boldsymbol{p}^k \cdot \boldsymbol{y}^l,$$
$$\boldsymbol{p}^l \cdot \boldsymbol{y}^l \geq \boldsymbol{p}^l \cdot \boldsymbol{y}^k.$$

Summing these inequalities, and rearranging terms, we get:

$$(\boldsymbol{p}^k - \boldsymbol{p}^l) \cdot (\boldsymbol{y}^k - \boldsymbol{y}^l) \geq 0.$$

This is the Law of Supply for an individual firm. The output quantities are increasing in own price (and input demands are decreasing in their price).

3 Maximizing an Objective Function

We will now make use of the utility representation of choice behavior, and turn the problem of describing choice behavior into a constrained optimization problem. This is an area where sharp mathematical tools (including calculus) yield sharp results.

3.1 Representing Continuous Rational Preferences

Since our domain for choices and preferences is for the most part $\mathbb{R}^L_+$, we define preferences on the positive orthant, i.e. $\boldsymbol{x} = (x_1, ..., x_L)$ such that $x_l \geq 0$ for all $l \in \{1, ..., L\}$.

Definition 3.1. An *upper contour set* of a rational preference relation $\succeq$ at $\boldsymbol{x} \in R^L$ is the set:

$$\mathcal{U}(\boldsymbol{x}; \succeq) := \{\boldsymbol{x}' | \boldsymbol{x}' \succeq \boldsymbol{x}\}.$$

A *lower contour set* of a rational preference relation $\succeq$ at $\boldsymbol{x} \in R^L$ is the set:

$$\mathcal{W}(\boldsymbol{x}; \succeq) := \{\boldsymbol{x}' | \boldsymbol{x} \succeq \boldsymbol{x}'\}.$$

Since $\succeq$ is complete, we have $\mathcal{U}(\boldsymbol{x}; \succeq) \cup \mathcal{W}(\boldsymbol{x}; \succeq) = \mathbb{R}^L_+$

Definition 3.2. A rational preference relation $\succeq$ is continuous if $\mathcal{U}(\boldsymbol{x}; \succeq)$ and $\mathcal{W}(\boldsymbol{x}; \succeq)$ are closed for all $\boldsymbol{x} \in \mathbb{R}^L$.

One of the key theorems in mathematical economics, due to Debreu (1954), shows that all continuous preference relations have a utility representation by a continuous utility function. The proof is a bit complicated and is really an exercise in mathematical analysis rather than economics so we skip it here. In almost all cases that we cover here (or in any other economics courses), we can safely assume that preferences are increasing in the sense that you cannot be hurt by having more of some good. We use the following notation for vector inequalities:

$$\boldsymbol{x} \geq \boldsymbol{x}' \iff x_l \geq x'_l \text{ for all } l \in \{1, ..., L\},$$

$$\boldsymbol{x} > \boldsymbol{x}' \iff x_l \geq x_l' \text{ for all } l \in \{1, ..., L\} \text{ and } \boldsymbol{x} \neq \boldsymbol{x}',$$

$$\boldsymbol{x} >> \boldsymbol{x}' \iff x_l > x_l' \text{ for all } l \in \{1, ..., L\}.$$

Definition 3.3. A rational preference relation is *monotonic* if $\boldsymbol{x} \geq \boldsymbol{x}'$ implies $\boldsymbol{x} \succeq \boldsymbol{x}'$. It is *strictly monotonic* if $\boldsymbol{x} > \boldsymbol{x}'$ implies $\boldsymbol{x} \succ \boldsymbol{x}'$.

With these definitions we can prove

Proposition 3.4. If $\succeq$ is a continuous strictly increasing preference relation on $\mathbb{R}_+^L$, then there exists a strictly increasing continuous utility function u on $\mathbb{R}^L$ representing $\succeq$.

Proof. Let $\mathbf{1} = (1, 1, ..., 1) \in \mathbb{R}^L$. Then $A = \{\alpha\mathbf{1} = (\alpha, ..., \alpha)\}$ is the diagonal of $\mathbb{R}_+^L$ as α ranges over $\mathbb{R}_+$. Consider an arbitrary $\boldsymbol{x} \in \mathbb{R}_+^L$. Since $\succeq$ is complete, $\alpha\mathbf{1} \in \mathcal{U}(\boldsymbol{x}; \succeq) \cup \mathcal{W}(\boldsymbol{x}; \succeq)$ for all $\alpha \in \mathbb{R}_+$. Since A is a closed and connected set and since both $\mathcal{U}(\boldsymbol{x}; \succeq)$ and $\mathcal{W}(\boldsymbol{x}; \succeq)$ are closed, there exists an $\alpha_x \in \mathbb{R}_+$ such that $\alpha_{\boldsymbol{x}}\mathbf{1} \in \mathcal{U}(\boldsymbol{x}; \succeq) \cap \mathcal{W}(\boldsymbol{x}; \succeq)$, i.e. $\alpha_{\boldsymbol{x}}\mathbf{1} \sim \boldsymbol{x}$. By strict monotonicity of $\succeq$, $\alpha_{\boldsymbol{x}}$ is unique. We claim first that $u(\boldsymbol{x}) = \alpha_{\boldsymbol{x}}$ represents $\succeq$. To see this, consider any $\boldsymbol{x} \succeq \boldsymbol{x}'$. We have by construction

$$\alpha_{\boldsymbol{x}}\mathbf{1} \sim \boldsymbol{x} \succeq \boldsymbol{x}' \sim \alpha_{\boldsymbol{x}'}\mathbf{1}.$$

By monotonicity,

$$\alpha_{\boldsymbol{x}}\mathbf{1} \succeq \alpha_{\boldsymbol{x}'}\mathbf{1} \iff \alpha_{\boldsymbol{x}} \geq \alpha_{\boldsymbol{x}'}.$$

Hence by our definition of u, we have $\boldsymbol{x} \succeq \boldsymbol{x}' \iff u(\boldsymbol{x}) \geq u(\boldsymbol{x}')$.

Finally, to see continuity, recall that a function $f : \mathbb{R}_+^L \to \mathbb{R}$ is continuous if and only if its upper and lower contour sets are closed. Since u represents $\succeq$, the assumption of continuous preferences guarantees that both $\{\boldsymbol{x}' | u(\boldsymbol{x}) \geq u(\boldsymbol{x}')\}$ and $\{\boldsymbol{x}' | u(\boldsymbol{x}') \geq u(\boldsymbol{x})\}$ are closed. □

3.1.1 Budget Sets and Walras' Law

In classical consumer theory, the set of available options for a decision maker is the same budget set that we saw in the subsection on revealed preference. For a strictly positive price vector $\boldsymbol{p} = (p_1, ..., p_L) >> 0$, and strictly positive income $w > 0$, we write:

$$B(\boldsymbol{p}, w) = \{\boldsymbol{x} \in \mathbb{R}^L_+ | \boldsymbol{p} \cdot \boldsymbol{x} = \sum_{l=1}^{L} p_l x_l \leq w\}.$$

Our first observation is that $B(\boldsymbol{p}, w)$ is homogenous of degree zero in $(\boldsymbol{p}, w)$, i.e. the budget set is unchanged if you multiply all prices and the income by the same positive number λ.

$$B(\boldsymbol{p}, w) = B(\lambda \boldsymbol{p}, \lambda w) \text{ for all } \lambda > 0.$$

Since the consumers preferences are defined on $\boldsymbol{x}$, the optimal choice does not change as the available set of options is the same. This is sometimes summarized by the statement that the consumer does not have money illusion:

$$\boldsymbol{x}(\boldsymbol{p}, w) = \boldsymbol{x}(\lambda \boldsymbol{p}, \lambda w) \text{ for all } \lambda > 0.$$

Euler's theorem for homogenous functions then gives:

Proposition 3.5. If a consumer has rational preferences on $\mathbb{R}^L_+$, then her demand function $\boldsymbol{x}(\boldsymbol{p}, w)$ satisfies:

$$\sum_{l=1}^{L} p_l \frac{\partial x_k(\boldsymbol{p}, w)}{\partial p_l} + w \frac{\partial x_k(\boldsymbol{p}, w)}{\partial w} = 0 \text{ for all } k.$$

Exercise 3.6. Show that the above expression can be written as:

$$\sum_{l=1}^{L} \epsilon_{lk} + \epsilon_{wk} = 0 \text{ for all } k,$$

where ϵ_{lk} is the price elasticity of good k with respect to price p_l, and ϵ_{wk} is the income elasticity of good k.

Note that $B(\boldsymbol{p}, w)$ is compact, i.e. closed (intersection of closed half-spaces) and bounded ($x_l \leq \frac{w}{p_l}$ for all l). Weierstrass' Theorem tells us that continuous functions attain their suprema on compact sets. In other words, the utility maximization problem:

$$\max_{\boldsymbol{x} \in B(\boldsymbol{p},w)} u(\boldsymbol{x}),$$

or

$$\max_{\boldsymbol{x}} u(\boldsymbol{x})$$
$$\text{subject to } \sum_{l=1}^{L} p_l x_l \leq w,$$

is well defined in the sense that it has a non-empty set of solutions. We summarize this as a theorem.

Theorem 3.7. The problem $\max_{\boldsymbol{x} \in B(\boldsymbol{p},w)} u(\boldsymbol{x})$ has a solution $\boldsymbol{x}(\boldsymbol{p}, w)$ called the *Walrasian or Marshallian demand correspondence* of the utility maximization problem.

The budget constraint $\boldsymbol{p} \cdot \boldsymbol{x} \leq w$ binds, i.e. is satisfied as an equality if the consumer has strictly monotonic preferences. The typical formulation in consumer theory is given in terms of slightly more general locally non-satiated preferences.

Definition 3.8. The rational preference relation $\succeq$ is *locally non-satiated* if for all $\boldsymbol{x}$ and all $\epsilon > 0$, there is another consumption vector $\boldsymbol{x}' \in B_\epsilon(\boldsymbol{x})$ such that $\boldsymbol{x}' \succ \boldsymbol{x}$.[8]

Exercise 3.9. Show that if u represents a locally non-satiated rational preference relation, then $\boldsymbol{p} \cdot \boldsymbol{x} = w$ for all $(\boldsymbol{p}, w)$.

[8] By $B_\epsilon(\boldsymbol{x})$, we denote the open ϵ-ball around $\boldsymbol{x}$, i.e.

$$B_{\epsilon(\boldsymbol{x})} = \left\{ \boldsymbol{y} \in \mathbb{R}^L | \sqrt{\sum_{l=1}^{L} (y_l - x_l)^2} < \epsilon \right\}$$

.

This observation goes under the name of Walras' Law. You will encounter it in competitive equilibrium models in Advanced Microeconomics 2. Again, as in the case of no money illusion, this simple observation has implications for the optimal demand. The following proposition results by totally differentiating the budget constraint, first with respect to income w, and then with respect to price p_k.

Proposition 3.10 (Engel and Cournot Aggregation)**.**
If $\boldsymbol{p} \cdot \boldsymbol{x}(\boldsymbol{p}, w) = w$ for all $(\boldsymbol{p}, w)$, then we have

1. $\sum_{l=1}^{L} p_l \frac{\partial x_l(\boldsymbol{p},w)}{\partial w} = 1$,
2. $\sum_{l=1}^{L} p_l \frac{\partial x_l(\boldsymbol{p},w)}{\partial p_k} + x_k(\boldsymbol{p}, w) = 0$ for all k.

Exercise 3.11. Show that the above expressions can be written in terms of elasticities as follows:

$$\sum_{l=1}^{L} s_l \epsilon_{wl} = 1,$$

$$\sum_{l=1}^{L} s_l \epsilon_{lk} + s_k = 0,$$

where $s_l := \frac{p_l x_l}{w}$ is the expenditure share of good l.

3.1.2 Additional Assumptions on Preferences

You may recall from earlier courses in microeconomics that indifference curves are normally drawn with a convex to origin shape. With a utility representation u for rational preferences, we could determine the shape by investigating the level curves $\{\boldsymbol{x} \in \mathbb{R}_+^L | u(\boldsymbol{x}) = \bar{u}\}$. If the utility function is twice differentiable, we could do this via Implicit Function Theorem, but this is quite cumbersome (just try it!). It is much more convenient to approach this directly through properties of $\succeq$.

Definition 3.12. A rational preference relation $\succeq$ is *convex* if $\boldsymbol{x} \succeq \boldsymbol{z}$, $\boldsymbol{y} \succeq \boldsymbol{z}$, and $\lambda \in [0, 1]$ imply $\lambda \boldsymbol{x} + (1 - \lambda)\boldsymbol{y} \succeq \boldsymbol{z}$. It is *strictly convex* if $\boldsymbol{x} \succeq \boldsymbol{z}$, $\boldsymbol{x} \neq \boldsymbol{y} \succeq \boldsymbol{z}$, and $\lambda \in (0, 1)$ imply $\lambda \boldsymbol{x} + (1 - \lambda)\boldsymbol{y} \succ \boldsymbol{z}$.

This may seem like an odd definition but if you look at two vectors $\boldsymbol{x} \sim \boldsymbol{y}$, they are located on the same indifference curve. For convex preferences, any weighted average of such $\boldsymbol{x}, \boldsymbol{y}$ is at least as good as $\boldsymbol{x}$ and $\boldsymbol{y}$. Strict convexity requires that all non-trivial weighted averages are strictly better than $\boldsymbol{x}$ and $\boldsymbol{y}$. Draw the pictures to see how this looks. If $\boldsymbol{x} \succeq \boldsymbol{z}$ and $\boldsymbol{y} \succeq \boldsymbol{z}$, then $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{U}(\boldsymbol{z}, \succeq)$. In other words, preferences are convex if all upper contour sets are convex sets.

Recall from Mathematics for Economists' (or MathCamp) the definition of a quasiconcave function:

Definition 3.13. A function f on a convex domain X is *quasiconcave* if for all $x, y \in X$ and all $\lambda \in [0,1]$,

$$f(\lambda x + (1-\lambda)y) \geq \min\{f(x), f(y)\}.$$

It is *strictly quasiconcave* if for all $x \neq y \in X$ and all $\lambda \in (0,1)$,

$$f(\lambda x + (1-\lambda)y) > \min\{f(x), f(y)\}.$$

Exercise 3.14. Suppose that u represents $\succeq$. Show that u is (strictly) quasiconcave if and only if $\succeq$ is (strictly) convex.

Mathematically the most important reason for assuming strictly convex preferences is that they result in single-valued solutions to the utility maximization problem.

Proposition 3.15. Suppose $\succeq$ is a continuous and strictly convex rational preference relation on a convex set B. Then the set of optimal choices from B is a singleton (i.e. consists of a single choice).

Proof. Since $\succeq$ is continuous, it has a continuous utility representation and Weierstrass' Theorem guarantees the existence of a $b^* \in B$ such that $b^* \succeq b$ for all $b \in B$.

We prove the claim by contradiction. Assume that $b^{**} \in B$ is another choice such that $b^{**} \succeq b$ for all $b \in B$. Since B is a convex set, $\frac{1}{2}b^* + \frac{1}{2}b^{**} \in B$.

By strict convexity of $\succeq$, we know that $\frac{1}{2}b^* + \frac{1}{2}b^{**} \succ b^*$ contradicting the optimality of b^*. Hence we conclude that there is a single b^* that dominates all other choices. $\square$

In order to use this result, we must show that the budget set $B(\boldsymbol{p}, w)$ is convex. If $\boldsymbol{p} \cdot \boldsymbol{x} \leq w$ and $\boldsymbol{p} \cdot \boldsymbol{x}' \leq w$, then

$$\boldsymbol{p} \cdot (\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{x}') \leq \lambda w + (1-\lambda) w = w,$$

and therefore $\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{x}' \in B(\boldsymbol{p}, w)$.

Now we know that under strictly convex preferences, i.e. strictly quasiconcave utility functions, the optimal demand $\boldsymbol{x}(\boldsymbol{p}, w) = \arg\max_{\boldsymbol{x} \in B(\boldsymbol{p}, w)} u(\boldsymbol{x})$ is indeed a function.

Exercise 3.16. Show that if $u : \mathbb{R}^L_+ \to \mathbb{R}$ is a (strictly) quasiconcave function and v is a strictly increasing function, then $v(u)$ is a (strictly) quasiconcave function.

You should already know that this result is true since u and $v(u)$ represent the same preferences and therefore the shape of the upper contour sets depends only on the underlying $\succeq$. Another exercise (familiar from the MathCamp) shows some sufficient conditions for quasiconcavity:

Exercise 3.17. If u is concave, then it is quasiconcave. By the previous exercise, this also implies that any strictly increasing function of a concave function is a quasiconcave function.

3.1.3 Some Details for the Interested

This may be a good time to look up Berge's Theorem of Maximum in Appendix M.K.2 of Mas-Colell, Whinston and Green. Since $B(\boldsymbol{p}, w)$ is a compact valued and continuous correspondence of $(\boldsymbol{p}, w)$ in the interior of $\mathbb{R}^{L+1}$, and $u(\boldsymbol{x})$ is continuous, the set of maximizers $\arg\max_{\boldsymbol{x} \in B(\boldsymbol{p}, w)} u(\boldsymbol{x})$ is a nonempty-valued upper-hemicontinuous correspondence. A single-valued upper-hemicontinuous correspondence is a continuous function.

As a result, we see that strictly convex preferences result in a continuous demand function in for interior $(\boldsymbol{p}, w)$.

Convex preferences result in a convex-valued demand correspondence, i.e.

$$\boldsymbol{x}, \boldsymbol{x}' \in \underset{\boldsymbol{x} \in B(\boldsymbol{p},w)}{\arg\max}\, u(\boldsymbol{x}) \implies \lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{x}' \in \underset{\boldsymbol{x} \in B(\boldsymbol{p},w)}{\arg\max}\, u(\boldsymbol{x}) \text{ for all } \lambda \in [0,1].$$

These properties of optimal demands will be useful in Advanced Microeconomics 2.

3.2 KKT Conditions for Utility and Profit Maximization

If we assume that the utility function is differentiable, we can make use of the KKT-conditions for maxima that should be familiar from Mathematics for Economists and from MathCamp.

3.2.1 Utility maximization problem (UMP)

A consumer allocates her budget of $w > 0$ to L goods. Her consumption vector is an element of the positive orthant of the L Euclidean space $X = \{\boldsymbol{x} \in \mathbb{R}_+^L\}$. We assume that the consumer has a continuous utility function $u(x)$ defined on X. Economic scarcity is present through the budget constraint:

$$\boldsymbol{p} \cdot \boldsymbol{x} \leq w \text{ or } \sum_{l=1}^{L} p_l x_l \leq w,$$

where $\boldsymbol{p} = (p_1, ..., p_L) >> 0$ is the vector of strictly positive prices for the goods. We can write this problem then as

Maximize

$$u(x_1, ..., x_L)$$

subject to

$$\sum_{l=1}^{L} p_l x_l \leq w,$$
$$x_l \geq 0 \text{ for all } l.$$

By writing the constraints in the equivalent form:

$$\sum_{l=1}^{L} p_l x_l - w \leq 0,$$
$$-x_i \leq 0 \text{ for all } l,$$

the problem is in the standard form that we always write for inequality constrained optimization problems.

Let's pause to see what we know about this problem already. A solution exists by Weierstrass' Theorem. If u is strictly increasing (as we usually assume) and quasiconcave, then the first order Kuhn-Tucker conditions are necessary and sufficient for optimum. In words, whenever we find a point satisfying the K-T conditions, we have solved the problem.

Let's turn our attention next to the Lagrangean and the K-T conditions:

$$\mathcal{L}(\boldsymbol{x}, \lambda) = u(\boldsymbol{x}) - \lambda_0 \left[\sum_{l=1}^{L} p_l x_l - w \right] + \sum_{l=1}^{L} \lambda_l x_l.$$

The first-order K-T conditions are:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}} = \frac{\partial u(\boldsymbol{x})}{\partial x_l} - \lambda_0 p_l + \lambda_l = 0 \text{ for all} l,$$
$$\lambda_0 \left[\sum_{l=1}^{L} p_l x_l - w \right] = 0,$$
$$\lambda_l x_l = 0 \text{ for all } l,$$
$$\sum_{l=1}^{L} p_l x_l - w \leq 0,$$
$$-x_l \leq 0 \text{ for all } l,$$
$$\lambda_l \geq 0 \;\; l \in \{0, 1, ..., L\}.$$

If the utility function has a strictly positive partial derivative for some x_i at the optimum, then the budget constraint must bind and $\lambda_0 > 0$. This follows immediately from the first line of the K-T conditions. For the other inequality constraints, consider the partial derivatives at $\boldsymbol{x} \in \mathbb{R}^L_+$ with $x_l \to 0$ for some l. If

$$\lim_{x_l \to 0} \frac{\partial u(\boldsymbol{x})}{\partial x_l} = \infty,$$

then we know again from the first line of the K-T conditions that at optimum $x_l > 0$. If this is true for all l, then we can ignore the non-negativity constraints and we are effectively back to a problem with a single equality constraint.

If $\frac{\partial u(\boldsymbol{x})}{\partial x_l} < \infty$ for $\boldsymbol{x} = (x_l, \boldsymbol{x}_{-l}) = (0, \boldsymbol{x}_{-l})$, then we must also consider corner solutions where $x_l = 0$ at optimum for some l. For interior solutions, we get from the first equation by eliminating λ the familiar condition:

$$\frac{\frac{\partial u}{\partial x_l}}{\frac{\partial u}{\partial x_k}} = \frac{p_l}{p_k}. \qquad (1)$$

This is of course the familiar requirement from intermediate microeconomics that at optimum, that $MRS_{x_l, x_k} = \frac{p_l}{p_k}$. Now we see that the same condition extends for many goods and the economic intuition is exactly the same. The price ration gives the marginal rate of transformation between the different goods and at an interior optimum, that rate must coincide with the marginal rate of substitution. In some cases, the functional form is such that the problem can be solved explicitly.

You have probably already some experience with solving explicitly constrained optimization problems. I include in the appendix to this section some examples of typical computations and you will be asked to solve some more in the problem sets. The main objective in this course is that you learn how to conceptualize the problems.

3.2.2 Firm's Profit Maximization Problem

In the most general setup, a firm is represented by a set of technologically feasible vectors $Y \subset \mathbb{R}^L$. Negative coordinates designate inputs and positive coordinates designate outputs.

The key assumptions are made on Y. Typically we make the following assumptions:

Definition 3.18. The production set Y of a competitive firm satisfies:

1. Y is non-empty and closed.
2. $Y \cap \mathbb{R}^L_+ = \{0\}$.
3. $\boldsymbol{y} \in Y$ and $\boldsymbol{y}' \leq \boldsymbol{y} \;\Rightarrow\; \boldsymbol{y}' \in Y$.
4. $\boldsymbol{y} \in Y \backslash \{0\} \;\Rightarrow\; -\boldsymbol{y} \notin Y$.

The first property makes the problem non-trivial and allows (potentially) the existence of profit-maximizing choices. The second says that you cannot produce something out of nothing (no free lunch) and inaction is possible. The third property allows for free disposal. The fourth rules out reversible production processes. Sometimes we also assume that Y is a convex set.

We can relate Y to familiar notions from intermediate microeconomics.

Definition 3.19. Y is said to have:

1. Decreasing returns to scale if: $\boldsymbol{y} \in Y \quad \Rightarrow \quad \alpha\boldsymbol{y} \in Y$ for all $\alpha \in [0,1]$.
2. Increasing returns to scale if: $\boldsymbol{y} \in Y \quad \Rightarrow \quad \alpha\boldsymbol{y} \in Y$ for all $\alpha \in [1,\infty)$.
3. Constant returns to scale if: $\boldsymbol{y} \in Y \quad \Rightarrow \quad \alpha\boldsymbol{y} \in Y$ for all α.

The firm is assumed to operate competitively, i.e. it takes a price vector $\boldsymbol{p} \in \mathbb{R}^L_+$ as given and maximizes profit:

$$\max_{\boldsymbol{y}\in Y} \boldsymbol{p} \cdot \boldsymbol{y}.$$

Since the constraint is written in a very general way, we cannot really write a Lagrangean for this problem in its usual form. We can nevertheless describe the boundary of the production set $\partial Y = \{\boldsymbol{y} \in Y | \neg \boldsymbol{y}' \in Y \text{ such that } \boldsymbol{y}' > \boldsymbol{y}\}$ by a function $F(\boldsymbol{y}) = 0$. This is called the transformation function for Y. For strictly positive price vectors, the firm's optimal choice is in ∂Y. Then the Lagrangean for the firm becomes:

$$\mathcal{L}(\boldsymbol{y}, \lambda) = \boldsymbol{p} \cdot \boldsymbol{y} - \lambda F(\boldsymbol{y}).$$

Exercise 3.20. Write the first-order conditions for optimal $\boldsymbol{y}$ and λ and interpret your result.

Since Y is unbounded, the existence of a solution is not obvious. For example, if Y satisfies constant returns to scale and some feasible vector results in a strictly positive profit, then no solution exist to the problem. We let $\boldsymbol{z}(p) = \arg\max_{\boldsymbol{y}\in Y} \boldsymbol{p} \cdot \boldsymbol{x}$ denote the (possibly empty) set of maximizers in this problem. The set of maximizers is convex if Y is a convex set.

Sometimes we can be more specific if we know that there is a single output y and K inputs $\boldsymbol{z} = (z_1, ..., z_K)$ that can be used to produce the output. In this case, we write:

$$y = f(z_1, ..., z_K),$$

with the understanding that y is the maximal quantity of output that can be produced from input vector $\boldsymbol{z}$. Profit maximization becomes then:

$$\max_{\boldsymbol{z}\in\mathbb{R}^K_+} pf(z_1,...,z_K) - \sum_{k=1}^{K} w_k z_k,$$

where $p > 0$ is the output price, and $\boldsymbol{w} = (w_1,...,w_K) >> 0$ is the vector of input prices. We have already seen in the part on revealed preference and profit that the firm's problem is easier to analyze than the consumer's problem. Notice that the constraint has been eliminated in this problem, and the Lagrangean coincides with the objective function.

We denote the solution $x(p,w_1,...,w_K)$, $\boldsymbol{z}(p,w_1,...,w_K)$ denote the maximizers to this problem. An interior solution to the problem exists if f satisfied the Inada conditions: $\lim_{z_k\to 0}\frac{\partial f(z_k,\boldsymbol{z}_{-k})}{\partial z_k} = \infty$, $\lim_{z_k\to\infty}\frac{\partial f(z_k,\boldsymbol{z}_{-k})}{\partial z_k} = 0$ for all $\boldsymbol{z}\in\mathbb{R}^K_{++}$. The maximizer is unique if f is a strictly quasiconcave function.

Exercise 3.21. Finnish forestry companies produce pulp using electricity as an input. They can also reverse the process and burn wood to produce electricity as an output. Draw a production set that allows for both of these production plans as optimal solutions depending on the market price of electricity.

Exercise 3.22. Draw the production set of a firm that must use a fixed amount $\overline{z}_1$ in order to produce any strictly positive amount of output q. Suppose that an additional input of Δ results in $\kappa\Delta$ units of output. Are there input and output prices such that the profit maximization problem of the firm has a solution?

3.3 Appendix: Examples of Computations

3.3.1 UMP with Constant Elasticity of Substitution

One of the most frequently used functional form in economics is the constant elasticity of substitution, CES function:

$$u(\boldsymbol{x}) = (a_1 x_1^\rho + \cdots + a_L x_L^\rho)^{\frac{1}{\rho}},$$

for $\rho < 1$. You can verify that u is quasiconcave e.g. by raising to power ρ and showing that the resulting function is concave (only non-zero terms in the Hessian are on the diagonal and negative).

We compute the marginal utility for each x_l:

$$\frac{\partial u}{\partial x_l} = \rho a_l x_l^{\rho-1} \frac{1}{\rho} (a_1 x_1^\rho + \cdots + a_L x_L^\rho)^{\frac{1}{\rho}}.$$

Note that since $\rho < 1$, we have $\frac{\partial u}{\partial x_l} > 0$, and

$$\lim_{x_l \to 0} \frac{\partial u}{\partial x_l} = \infty.$$

Since the feasible set is convex and the objective function is quasiconcave with a non-vanishing derivative, the first order conditions are necessary and sufficient for optimum. Since the marginal utility is unbounded at the boundary, we know that we have an interior solution and that the budget constraint is binding. Hence the KKT conditions require simply that for all i, k:

$$\frac{\frac{\partial u}{\partial x_l}}{\frac{\partial u}{\partial x_k}} = \frac{p_l}{p_k},$$

and the budget constraint holds as an equality:

$$\sum_{l=1}^{L} p_l x_l = 0.$$

Hence we have that

$$\frac{a_1 x_1^{\rho-1}}{a_k x_k^{\rho-1}} = \frac{p_1}{p_k},$$

or

$$\frac{x_1}{x_k} = (\frac{a_k p_1}{a_1 p_k})^{\frac{1}{\rho-1}},$$

or

$$x_k = x_1 (\frac{a_k p_1}{a_1 p_k})^{\frac{1}{1-\rho}}. \tag{2}$$

Substituting into the budget constraint, we get:

$$p_1 x_1 + \sum_{l=2}^{L} p_l x_1 (\frac{a_l p_1}{a_1 p_l})^{\frac{1}{1-\rho}} = w.$$

We can solve for x_1 to get

$$x_1 = \frac{w}{p_1 + \sum_{l=2}^{L} p_l (\frac{a_l p_1}{a_1 p_l})^{\frac{1}{1-\rho}}}.$$

Substituting into (2), we can solve the other x_j.

To get a bit nicer expression, let $r = \frac{\rho}{\rho-1}$ and assume that $a_l = 1$ for all l. Then we have for each j:

$$x_j = \frac{w p_j^{r-1}}{\sum_{l=1}^{L} p_l^r}.$$

In this case, we are able to solve the optimal demands as explicit functions of the exogenous variables. We call the optimal solutions to the utility maximization problem the Marshallian demands. You have probably seen these demand functions in models of monopolistic competition as needed in growth theory, international trade and industrial organization.

If you want to understand where the name constant elasticity of substitution comes from, you should note that:

$$\frac{x_l}{x_k} = (\frac{p_l}{p_k})^{\frac{1}{\rho-1}} (\frac{a_k}{a_l})^{\frac{1}{\rho-1}}.$$

Hence a small percentage change in the price ratio between any two goods induces the same percentage change in the ratio of the optimal consumptions. The size of this change is given by $\frac{1}{\rho-1}$ and hence ρ measures the elasticity of substitution between any two goods. The higher, ρ, the higher the substitution away from a good when its price increases.

You should consider the comparative statics of the optimal demands in prices and income. In other words, compute the partial derivatives $\frac{\partial x_l(\boldsymbol{p},w)}{\partial p_l}$, $\frac{\partial x_l(\boldsymbol{p},w)}{\partial p_j}$ and $\frac{\partial x_i(\boldsymbol{p},w)}{\partial w}$. For example, when does the demand for good i increase in the price of another good p_j?

3.3.2 Cobb-Douglas utility function

Let's look at some special cases. By l'Hôpital's rule, the CES -function converges to the *Cobb-Douglas utility function* $u(\boldsymbol{x}) = x_1^{\alpha_1} \dots x_L^{\alpha_L}$ as $\rho \to 0$.

If we just substitute $\rho = 0$ into the optimal demand, we get

$$x_l = \frac{\alpha_l w}{p_l(\sum_{l=1}^{L} \alpha_l)}.$$

For the Cobb-Douglas utility function, you get the result that the expenditure share $\frac{p_l x_l}{w}$ on each good is equal to $\frac{\alpha_l}{(\sum_{l=1}^{L} \alpha_l)}$. In this case, the consumer's expenditure share does not depend on her wealth. In other words, rich and poor consumers use the same fraction of their income on food, clothing, yachts etc. This is clearly not a very good description of reality.By equation (2), you can see that CES -functions do not offer that much help either. The expenditure shares are still constant in wealth (even though they depend now on the entire price vector).

3.3.3 Stone-Geary utility function

One way to get more realistic consumption patters is to define the utility function for consumptions above a level needed for subsistence. Let $\underline{\boldsymbol{x}} = (\underline{x_1}, ..., \underline{x_L})$ be the levels of each good needed for survival and assume that

$w \geq \boldsymbol{p} \cdot \underline{\boldsymbol{x}}$. The utility function for $\boldsymbol{x} \in \mathbb{R}^L$ such that $x_l \geq \underline{x_l}$ is of Cobb-Douglas -like form:

$$u(\boldsymbol{x}) = (x_1 - \underline{x_1})^{\alpha_1} \dots (x_L - \underline{x_L})^{\alpha_L},$$

where $0 < \alpha_l < 1$ for all l and $\sum_{l=1}^{L} \alpha_l = 1$. This utility specification is known as the *Stone-Geary utility function*. Notice that the marginal utility for good i is infinite if $x_l = \underline{x_l}$ and that the utility function is strictly increasing in all of its components. Hence we still have an interior solution and the budget constraint binds.

We get as above:

$$\frac{\frac{\partial u(\boldsymbol{x})}{\partial x_l}}{\frac{\partial u(\boldsymbol{x})}{\partial x_k}} = \frac{\alpha_l(x_k - \underline{x_k})}{\alpha_k(x_l - \underline{x_l})} = \frac{p_l}{p_k} \text{ for all } l, k,$$

and

$$\sum_{l=1}^{L} p_l x_l = w.$$

We get that

$$x_l - \underline{x_l} = \frac{\alpha_l p_1}{\alpha_1 p_l}(x_1 - \underline{x_1}) \text{ for all } l. \tag{3}$$

Multiplying both sides by p_l and summing over l gives:

$$\sum_{l=1}^{L} p_i(x_l - \underline{x_l}) = \frac{p_1 \sum_{l=1}^{L} \alpha_l}{\alpha_1}(x_1 - \underline{x_1}).$$

So we can solve:

$$x_1 - \underline{x_1} = \frac{\alpha_1(w - \sum_{l=1}^{L} p_l \underline{x_l})}{p_1},$$

where we used the budget constraint $\sum_{l=1}^{L} p_l x_l = w$ and $\sum_{l=1}^{L} \alpha_l = 1$

By (3), we see that

$$x_k - \underline{x_k} = \frac{\alpha_k(w - \sum_{l=1}^{L} p_l \underline{x_l})}{p_k}.$$

Now you can see that the consumer uses a constant fraction of her excess income (above what is needed for the necessities $\underline{\boldsymbol{x}}$) in constant shares given by the α_i. Since the poor have less excess wealth, their consumption fractions are closer to the ones given by the subsistence levels $\beta_l := \frac{\underline{x_l}}{\sum_l \underline{x_l}}$. Hence the richest spend fractions α_l on good l and the poorest spend β_l.

3.4 Comparative Statics of Optimal Choices

When solving a maximization problem, the end result is in the best case scenario a vector in the feasible set. As such, it does not tell us much about the economic forces at play. We are much more interested in knowing what happens to the solution as the parameters of the problem change: How does your consumption choices change if the price of gasoline goes up? What happens to your savings decisions if interest rates go up? What happens to your demand for insurance if uncertainty about future events increases? Answering these questions is called the comparative statics of the model.

3.4.1 Brute Force: IFT

If we can solve our maximization problem, say $\boldsymbol{x}(\boldsymbol{p}, w)$ explicitly (i.e. in closed form), we can answer such questions by just differentiating the solution with respect to the parameters. Unfortunately, this is rarely the case. All you have after solving the model is a set of implicit equations from the KKT-conditions that you know hold at the optimum. In the best case scenario, (e.g. by having strictly quasiconcave objective and quasiconvex constraints), you may know that the solution is uniquely pinned down by the first-order conditions, but how do you proceed from there?

One possibility is brute force. If we have sufficient differentiability in our objective functions and constraints, we can apply the Implicit Function Theorem around the solution. Let's try this for some relatively simple problems.

Example 3.23. Consider the following problem (with interpretations to saving, choice under uncertainty etc.)

$$max_{x_1,x_2}\pi_1 u(x_1) + \pi_2 v(x_2)$$

$$\text{s.t. } p_1x_1 + p_2x_2 - w = 0.$$

If the functions are nice enough, say $u(x) = v(x) = \ln x$, comparative statics follow immediately from the solutions

$$x_1 = \frac{\pi_1 w}{p_1(\pi_1 + \pi_2)}, x_2 = \frac{\pi_2 w}{p_2(\pi_1 + \pi_2)}$$

If u and v are not so nice, we need to look at the KKT first-order conditions:

$$\pi_1 u' - \lambda p_1 = 0,$$

$$\pi_2 v' - \lambda p_2 = 0,$$

$$p_1x_1 + p_2x_2 - w = 0.$$

Then we can use implicit function theorem to get for the effects of a change in p_1:

$$\begin{pmatrix} \pi_1 u'' & 0 & -p_1 \\ 0 & \pi_2 v'' & -p_2 \\ p_1 & p_2 & 0 \end{pmatrix} \begin{pmatrix} dx_1 \\ dx_2 \\ d\lambda \end{pmatrix} = \begin{pmatrix} \lambda \\ 0 \\ -x_1 \end{pmatrix} dp_1.$$

Solving for $\frac{dx_1}{dp_1}$ and $\frac{dx_2}{dp_1}$ gives (by Cramer's rule):

$$\frac{dx_1}{dp_1} = \frac{1}{|H|}(\lambda p_2^2 - p_1 x_1 \pi_2 v''),$$

$$\frac{dx_2}{dp_1} = \frac{1}{|H|}(p_2(-\lambda p_1 - x_1 \pi_1 u''),$$

where $|H|$ is the determinant of the Hessian of the Lagrangean, and by second-order conditions for local maximum, $|H| \leq 0$. For changes in π_2:

$$\frac{dx_1}{d\pi_2} = \frac{1}{|H|}(p_1 p_2 v'),$$

Obviously, you can write similar expressions for more general $u(x_1, x_2)$, but now you will have also cross-partials $u_{12}(x_1, x_2)$in the expressions. It gets a bit messy as you can see if you work out the following recursive utility formulation with utility function $U(x_1, u(x_2))$.

Exercise 3.24. Consider the following consumer problem:

$$\max_{x_1, x_2} = U(x_1, u(x_2))$$

$$\text{s.t. } p_1 x_1 + p_2 x_2 - w = 0.$$

Find the comparative statics of x_2 in w. (Hint: assume interior solution and substitute $x_1 = \frac{w - p_2 x_2}{p_1}$ to get to an unconstrained problem in a single variable).

As you can see, this approach results quite quickly in long strings of derivatives of different orders, and it is quite hard to sign the comparative statics. We look next at an alternative that is deceptively simple, and works sometimes in settings where implicit function theorem does not apply (e.g. discrete choice variables).

3.4.2 Monotone Comparative Statics

We outline here an approach to monotone comparative statics that does not hinge on differentiability, but on properties or orderings on the decision variables and the parameters of the model. The idea is to look again at parametrized optimization problems.

The simplest setting for such considerations is with a family of functions $\{f(x, \theta)\}_{\theta \in \Theta}$, where $x \in X$ is the choice variable for some $X \subset \mathbb{R}$, and $\theta \in \Theta$ is a real-valued parameter for the problem: $\Theta \subset \mathbb{R}$.

The basic question here is to determine which properties of f ensure that any selection from the set of optimal choices $x(\theta) \in X(\theta)$ is a monotone function. We focus on the case where $x(\theta)$ is increasing (or non-decreasing).

Here are the two key notions for these notes: single-crossing and strictly increasing differences.

Definition 3.25 (Single Crossing)**.** A function $g : \Theta \to \mathbb{R}$ is *single crossing* if for all $\theta^{''} > \theta^{'}$,

$$g(\theta^{'}) \geq (>)0 \implies g(\theta^{''}) \geq (>)0.$$

Definition 3.26 (Strictly Single Crossing)**.** A function $g(s)$ is *strictly single crossing* if for all $\theta'' > \theta'$,

$$g(\theta') \geq 0 \implies g(\theta^{''}) > 0.$$

Definition 3.27 (Single Crossing Differences)**.** A family of functions $\{f(\cdot, \theta)\}_{\theta \in \Theta}$ has *(strictly) single crossing differences* if for all $x'' > x'$, the function

$$\delta(\theta) := f(x'', \theta) - f(x', \theta)$$

is (strictly) single crossing.

Notice that these are ordinal properties. It is a good exercise to show that if $\{f(\cdot, \theta)\}_{\theta \in \Theta}$ has single crossing difference and $h(x, \theta)$ is increasing in x, then $\{g(\cdot, \theta)\}_{\theta \in \Theta}$ where $g(x, \theta) := f(h(x, \theta), \theta)$ has also single crossing differences.

Definition 3.28 (Strictly Increasing Differences)**.** A family of functions $\{f(\cdot, \theta)\}_{\theta \in \Theta}$ has *strictly increasing differences* (SID) if $\delta(\theta)$ is strictly increasing. We also say that $f(x, \theta)$ has SID.

Remark. It is easy to see that if $f(x, \theta)$ has SID, then it also satisfies the strict single crossing property and single crossing differences, and if f is differentiable in θ, , then $f_\theta(x, \theta)$ is increasing in x. If it is twice differentiable, then a necessary and sufficient condition for (strictly) increasing differences is that $f_{\theta x}(x, \theta) \geq (>)0$.

Define next an order on families of sets as follows. Let $Y(\theta) \subset \mathbb{R}$ be sets parametrized by θ.

Definition 3.29. $Y(\theta)$ is increasing (in the strong set order) if for $\theta' > \theta$,

$$x' \in Y(\theta') \text{ and } x \in Y(\theta) \ \Rightarrow \max\{x', x\} \in Y(\theta') \text{ and } \min\{x', x\} \in Y(\theta).$$

The main tool in monotone comparative statics analysis is the following theorem due to Milgrom and Shannon (1994).

Theorem 3.30 (Milgrom and Shannon)**.** The family $\{f(\cdot, \theta)\}_{\theta \in \Theta}$ has single crossing differences if and only if $\arg\max_{x \in Y} f(x, \theta)$ is increasing in θ for all $Y \subseteq X$.

Hence if $f(x, \theta)$ has SID, the set of maximizers is increasing (in the strong set order). If f is strictly quasiconcave in x, then the unique maximizer is increasing in θ. The proof of this Theorem is left as an exercise.

This approach is indeed quite straightforward and does not involve complicated computations as much as implicit function theorem. At the same time, there are some shortcomings. You may note that up to this point, we have assumed $X \subset \mathbb{R}$. This guarantees that $\min\{x, x'\} \in X$ and $\max\{x, x'\} \in X$ if $x, x' \in X$. This is clearly not necessarily true if $X \subset \mathbb{R}^L$ for $L > 1$. By defining $\max\{x, x'\}$ and $\min\{x, x'\}$ for all $x, x' \in X$, we are defining a partial order on X. A partially ordered set X is called a *lattice* if for all $x, x' \in X$, $\min\{x, x'\}$ and $\max\{x, x'\} \in X$. We typically denote the minimum and maximum of two vectors by $x \wedge x'$ and $x \vee x'$. For $X \subset \mathbb{R}^L$, we define $\boldsymbol{x} \wedge \boldsymbol{x}' = (\min\{x_1, x_1'\}, ..., \min\{x_L, x_L'\})$, and $\boldsymbol{x} \vee \boldsymbol{x}' = (\max\{x_1, x_1'\}, ..., \max\{x_L, x_L'\})$, i.e. the componentwise minimum and maximum of the two vectors.

Definition 3.31. We say that a function $f(\boldsymbol{x}, \theta)$ on X, Θ is *supermodular* in $\boldsymbol{x}$ if X is a lattice, and for all $\boldsymbol{x}, \boldsymbol{x}' \in X$, and for all $\theta \in \Theta$,

$$f(\boldsymbol{x} \vee \boldsymbol{x}', \theta) + f(\boldsymbol{x} \wedge \boldsymbol{x}', \theta) \geq f(\boldsymbol{x}, \theta) + f(\boldsymbol{x}', \theta).$$

Exercise 3.32. Show that if $f(\boldsymbol{x}, \theta)$ is twice differentiable in $\boldsymbol{x}$, then f is supermodular in $\boldsymbol{x}$ if and only if $\frac{\partial^2 f(\boldsymbol{x}, \theta)}{\partial x_i \partial x_j} \geq 0$ for all $i \neq j \in \{1, ..., L\}$.

If X is a lattice and if Θ is a partially ordered set,[9] we define increasing differences as before.

Definition 3.33. Suppose that X is a lattice and Θ is a partially ordered set. A function $f : X \times \Theta \to \mathbb{R}$ has *increasing differences in* $(\boldsymbol{x};\theta)$ if for all $\boldsymbol{x}' > \boldsymbol{x}$ and $\theta' > \theta$,

$$f(\boldsymbol{x}', \theta') - f(\boldsymbol{x}, \theta') \geq f(\boldsymbol{x}', \theta) - f(\boldsymbol{x}, \theta).$$

One of the most fundamental theorems in monotone comparative statics is the following due to Topkis (1978).

Theorem 3.34 (Topkis)**.** Suppose X is a lattice, Θ is a partially ordered set, and $f : X \times \Theta \to \mathbb{R}$. If f is supermodular in $\boldsymbol{x}$, and has increasing differences in $(\boldsymbol{x};\theta)$, then $\arg\max_{\boldsymbol{x}\in X} f(\boldsymbol{x},\theta)$ is monotone nondecreasing in θ (in the strong set order).

Proof. Let $X^*(\theta) = \arg\max_{x\in X} f(x,\theta)$. Fix any $\theta' \geq \theta$, and $\boldsymbol{x} \in X^*(\theta)$ and $\boldsymbol{x}' \in X^*(\theta')$. Note that

$$\begin{aligned}
0 &\leq f(\boldsymbol{x},\theta) - f(\boldsymbol{x}\wedge\boldsymbol{x}',\theta) \text{ since } \boldsymbol{x} \in X^*(\theta),\\
&\leq f(\boldsymbol{x}\vee\boldsymbol{x}',\theta) - f(\boldsymbol{x}',\theta) \text{ by supermodularity,}\\
&\leq f(\boldsymbol{x}\vee\boldsymbol{x}',\theta') - f(\boldsymbol{x}',\theta') \text{ by increasing differences ,}\\
&\leq 0 \text{ since } \boldsymbol{x}' \in X^*(\theta').
\end{aligned}$$

Thus $f(\boldsymbol{x},\theta) = f(\boldsymbol{x}\wedge\boldsymbol{x}',\theta)$ and $f(\boldsymbol{x}\vee\boldsymbol{x}',\theta) = f(\boldsymbol{x}',\theta)$, which implies $\boldsymbol{x}\wedge\boldsymbol{x}' \in X^*(\theta)$ and $\boldsymbol{x}\vee\boldsymbol{x}' \in X^*(\theta')$. □

If $f(\boldsymbol{x},\theta)$ is strictly quasiconcave in $\boldsymbol{x}$, then the unique maximizer is increasing in θ.

Monotone comparative statics often helps you to avoid the tedious calculations of implicit function theorem. Verifying strictly increasing differences (and supermodularity) is sometimes quite easy. You should also

[9]This means that there is a transitive, reflexive and antisymmetric binary relation $\geq$ defined on Θ

note that X and Θ may be discrete sets (as long as the former is a lattice and the latter is partially ordered). Unfortunately not all interesting feasible sets are lattices in economics. The most obvious example is the budget set $B(\boldsymbol{p}, w)$. Monotone comparative statics is still extremely useful for many decision problems under uncertainty, firm's problems and in the theory of mechanism design. We end with some examples of this method.

Example 3.35. Consider the effects of an increase in the market size on monopoly price. Each consumer in this market has an inverse demand given by $P(q)$. If the number of customers N is exogenous, then firm's problem is to choose a quantity per customer q to maximize:

$$\pi(q, N) = NqP(q) - C(Nq).$$

Without some restrictions on demand, we cannot conclude that π has increasing differences in $(q; N)$. However, note that for$N > 0$, we have

$$\arg\max_{q\geq 0} \pi(q, N) = \arg\max_{q\geq 0} \pi(q, N) = \arg\max_{q\geq 0} qP(q) - C(Nq).$$

Thus, $q^*(N)$ is nondecreasing in N if $-C(Nq)/N$ has increasing differences in $(q; N)$. Assuming that C is twice continuously differentiable (this assumption is not required for the conclusion but makes the characterization of increasing differences simpler),

$$\frac{d^2(-C(Nq)/N)}{dNdq} = \frac{d(-C'(Nq))}{dN} = -qC''(Nq).$$

Thus $q^*(N)$ is nondecreasing in N if C is concave, and $q^*(N)$ is nonincreasing in N if C is convex. No restrictions on the inverse demand function are required for this conclusion. It is important to note that taking a transformation that depends on the choice variable x would alter the set of maximizers. For example, dividing by q in the previous example would change the solution set and could therefore yield incorrect conclusions. Be careful when applying this method that your transformation only involves parameters and not choice variables.

Exercise 3.36. Consider a two-period consumption-savings problem. The suppose the individual has the following utility function for $(c_1, c_2) \in \mathbb{R}^2_+$:

$$U(c_1, c_2) = u(c_1) + \beta u(c_2)$$

where $\beta \in (0,1)$ and $u : \mathbb{R}_+ \to \mathbb{R}$ is strictly increasing. The individual has initial wealth w at the start of period 1 and can save between the two periods at a (deterministic) gross interest rate of $(1 + r) > 0$.

1. Suppose you know that u is twice continuously differentiable and concave, $u'' \leq 0$. What conclusions can be made about how the optimal c_1 and c_2 change with wealth w? (Hint: Substitute out one of the consumptions using the budget constraint).

2. In the previous parts of the problem, is it possible to have a selection $c_1^*(w)$ from the optimal consumption choices in period 1 that strictly decreases in w at some wealth levels, so $w' > w$ and $c_1^*(w') < c_1^*(w)$? Either provide an example where this can happen or prove that it cannot happen. Also, if your answer is that $c_1^*(w)$ can strictly decrease in w under the assumptions given in this problem, what change in the assumptions would ensure that every selection $c_1^*(w')$ from the solution is monotonically nondecreasing in w?

3.5 Value Function and Envelope Theorem

In this last subsection, we will look at the value functions related to maximization and minimization problems. The value function to an optimization problem is obtained simply by substituting a solution of the problem to the objective function. Let's see how this goes through an example

Example 3.37. Consider the UMP $\max_{\boldsymbol{x} \in B(\boldsymbol{p},w)} u(\boldsymbol{x})$ with a strictly quasi-concave u so that the solution $\boldsymbol{x}(\boldsymbol{p}, w)$ is single-valued for all $(\boldsymbol{p}, w)$. The *value function* of the problem is then $v(\boldsymbol{p}, w) := u(\boldsymbol{x}(\boldsymbol{p}, w))$. This value function is called the *indirect utility function* of the problem, and it records the maximal utility level attainable at the budget set $B(\boldsymbol{p}, w)$. Note that the numerical value is not really important since the utility function itself is only ordinal representation of the preferences.

In general, we can consider any parametric family of optimization problems, $\max_{x \in X(\alpha)} f(x;\alpha)$, where α is the parameter (possibly vector) of interest. If $x(\alpha)$ is a solution to the problem, then $V(\alpha) = f(x(\alpha);\alpha)$ is the value function of the problem.

In the next section, we shall see that the value functions of optimization problems contain almost all relevant information for the problems. Before that, we discuss one of the most important results in mathematical economics: the envelope theorem. The envelope theorem gives a concrete interpretation to Lagrange multipliers and it is an extremely useful tool for comparative statics. Assume that the objective function u and the constraint functions g_k are differentiable in $(\boldsymbol{x}, \alpha)$.

Theorem 3.38 (Envelope Theorem)**.** If $V(\alpha)$ is the value function of a constrained optimization problem with an objective function $u(\boldsymbol{x}, \alpha)$ on $\mathbb{R}^L$ depending on a parameter $\alpha \in \mathbb{R}$ with K binding constraints and Lagrangean

$$\mathcal{L}(\boldsymbol{x}, \lambda; \alpha) = u(\boldsymbol{x}; \alpha) - \sum_{k=1}^{K} \lambda_k g_k(\boldsymbol{x}; \alpha),$$

then:

$$V'(\alpha) = \frac{\partial u(\boldsymbol{x}^*, \alpha)}{\partial \alpha} - \sum_{k=1}^{K} \lambda_k^* \frac{\partial g_k(\boldsymbol{x}^*; \alpha)}{\partial \alpha}.$$

The proof is in the Appendix to this section.

Note that for the case of unconstrained optimization problems, we get simply:

$$V'(\alpha) = \frac{\partial u(\boldsymbol{x}; \alpha)}{\partial \alpha}.$$

In fact, this result is much more general, and really only requires differentiability of u in α. The following is a slightly simplified version of Milgrom and Segal (2002), and it is valid for $\alpha \in [\underline{\alpha}, \overline{\alpha}]$.

Theorem 3.39 (Milgrom and Segal, 2002)**.** Assume that

- $f(x;\alpha)$ is differentiable in α with a uniformly bounded derivative

$$\frac{\partial f(x;\alpha)}{\partial \alpha} \leq K < \infty \text{ for all } x \in X,$$

- The set of optimizers $X^*(\alpha) \neq \emptyset$ for all α.

Then $V(\alpha)$ is absolutely continuous, and for any selection $x^*(\alpha)$ from $X^*(\alpha)$,

$$V'(\alpha) = \frac{\partial f(x^*;\alpha)}{\partial \alpha}$$

for almost every α, and therefore

$$V(\alpha) = V(\underline{\alpha}) + \int_{\underline{\alpha}}^{\alpha} \frac{\partial f(x^*(s);s)}{\partial \alpha} ds.$$

You will see versions of envelope theorem in this course, and later also in macroeconomics (dynamic programming) and in economics of uncertainty (Advanced Microeconomics 4). Let's list some important economic instances where value functions play a key role.

1. Indirect utility function $v(\boldsymbol{p}, w) = \max_{\boldsymbol{x}\in B(\boldsymbol{p},w)} u(\boldsymbol{x})$.

2. Profit function $\pi(\boldsymbol{p}) = \max_{\boldsymbol{y}\in Y} \boldsymbol{p}\cdot\boldsymbol{y}$.

3. Cost function $c(\boldsymbol{w}, q) = \min_{\boldsymbol{z}:f(\boldsymbol{z})\geq q} \boldsymbol{w}\cdot\boldsymbol{z}$.

4. Expenditure function: $e(\boldsymbol{p}, \overline{u}) = \min_{\boldsymbol{x}:u(\boldsymbol{x})\geq \overline{u}} \boldsymbol{p}\cdot\boldsymbol{x}$.

5. Value of information: $V(\boldsymbol{\pi}) = \max_{x\in X} \sum_{i=1}^{I} \pi_i u(x, \omega_i)$.

6. Value with types: $V(\theta) = \max_{(x,p)\in\{(x_i,p_i)\}_{i=i}^{I}} v(x,\theta) - p$.

You may note that if we were really pedantic, the indirect utility function should also be indexed by the underlying utility function u, the profit function by Y, etc. The notation above suggests that we will keep these fixed throughout and we have listed only the parameters of interest to our comparative statics. We have already discussed the utility and profit maximization problems. The cost function outputs the smallest cost at which you can reach the output target q if input cost vector is $\boldsymbol{w}$ and the production function is f. Expenditure function is the cost function of the consumer to reach utility level $\overline{u}$ at prices $\boldsymbol{p}$. We'll talk more about the value of information in Section 5 of this course. You can think of it as the maximal expected utility that you can reach in a decision problem under uncertainty if your belief over the states of nature is given by the probability vector $\boldsymbol{\pi}$. The last one you will see in Advanced Microeconomics 4. It gives the maximal value to a buyer with preference parameter θ if she can choose from a menu $\{(x_i, p_i)\}_{i=i}^{I}$ of I different options x_i offered at associated prices p_i.

As an aside, you should be able to say something about the comparative statics about the optimal solutions in some of these problems in light of our previous section on monotone comparative statics. Just see if you have strictly increasing differences in the choice variable and the parameter (and if the choice is over vectors, check supermodularity).

Here is a key proposition on the shape of value functions.

Proposition 3.40. If the choice set X does not depend on $\alpha \in A$, and A is a convex set, then:

1. $V(\alpha) = \max_{x\in X} u(x;\alpha)$ is convex if $u(x;\alpha)$ is convex in α.

2. $V(\alpha) = \min_{x\in X} u(x;\alpha)$ is concave if $u(x;\alpha)$ is concave in α.

Proof. We prove the claim in the first case. The second case is analogous and left as an exercise. We need to show that for all $\alpha, \alpha' \in A$ and for all $\lambda \in [0,1]$:

$$\lambda V(\alpha) + (1-\lambda)V(\alpha') \geq V(\lambda\alpha + (1-\lambda)\alpha').$$

Let x solve $\max_{x\in X} u(x;\alpha)$, x' solve $\max_{x\in X} u(x;\alpha')$, and x^λ solve $\max_{x\in X} u(x;\lambda\alpha + (1-\lambda)\alpha')$. Since u is convex in α by assumption, we have:

$$\lambda u(x^\lambda;\alpha) + (1-\lambda)u(x^\lambda;\alpha') \geq u(x^\lambda;\lambda\alpha + (1-\lambda)\alpha').$$

By definition,

$$u(x;\alpha) \geq u(x^\lambda;\alpha) \text{ and } u(x';\alpha) \geq u(x^\lambda;\alpha').$$

Therefore we have:

$$\lambda V(\alpha) + (1-\lambda)V(\alpha') = \lambda u(x;\alpha) + (1-\lambda)u(x';\alpha')$$
$$\geq u(x^\lambda; \lambda\alpha + (1-\lambda)\alpha') = V(\lambda\alpha + (1-\lambda)\alpha').$$

□

Since linear functions are concave and convex, this tells us immediately that the profit function and the value of information are convex functions whereas the expenditure function and the cost function are concave. We cannot say anything at this stage about the indirect value function since there the choice set depends on parameters. The value of information is convex in π, and value with types is convex if $v(x,\theta)$ is convex in θ.

Let's end this subsection with two results on indirect utility function. The first is an application of the envelope theorem.

Proposition 3.41 (Roy's Lemma)**.** Suppose $v(\boldsymbol{p}, w)$ is differentiable. Then for all $l \in \{1, ..., L\}$, we have:

$$x_l(\boldsymbol{p}, w) = -\frac{\frac{\partial v(\boldsymbol{p},w)}{\partial p_l}}{\frac{\partial v(\boldsymbol{p},w)}{\partial w}}.$$

Proof. The Lagrangean of the utility maximization problem is:

$$\mathcal{L} = u(\boldsymbol{x}) - \lambda(\sum_{l=1}^{L} p_l x_l - w).$$

By envelope theorem,

$$\frac{\partial v(\boldsymbol{p}, w)}{\partial p_l} = -\lambda x_l,$$
$$\frac{\partial v(\boldsymbol{p}, w)}{\partial w} = \lambda.$$

Therefore the result follows by division. □

Proposition 3.42. The indirect utility function $v(\boldsymbol{p}, w)$ is quasiconvex in $(\boldsymbol{p}, w)$

Proof. To prove this, we must show that $v(\lambda\boldsymbol{p} + (1-\lambda)\boldsymbol{p}', \lambda w + (1-\lambda)w') \leq \max\{v(\boldsymbol{p}, w), v(\boldsymbol{p}', w')\}$. Let $\boldsymbol{x}^\lambda$ denote the optimal choice from $B(\lambda\boldsymbol{p} + (1-\lambda)\boldsymbol{p}', \lambda w + (1-\lambda)w')$.

Since $\boldsymbol{x}^\lambda \in B(\lambda\boldsymbol{p} + (1-\lambda)\boldsymbol{p}', \lambda w + (1-\lambda)w')$, either $\boldsymbol{x}^\lambda \in B(\boldsymbol{p}, w)$ or $\boldsymbol{x}^\lambda \in B(\boldsymbol{p}', w')$, and the claim follows. □

3.5.1 Appendix: Proof of Differentiable Envelope Theorem

Envelope Theorem.

$$V'(\alpha) = \frac{\partial u(\boldsymbol{x}^*, \alpha)}{\partial \alpha} + \sum_{i=1}^{L} \frac{\partial u(\boldsymbol{x}^*, \alpha)}{\partial x_i} x_i'(\alpha).$$

The first-order condition w.r.t. x_i gives:

$$\frac{\partial u(\boldsymbol{x}^*, \alpha)}{\partial x_i} - \sum_{k=1}^{K} \lambda_k^* \frac{\partial g_k(\boldsymbol{x}^*, \alpha)}{\partial x_i} = 0,$$

so that

$$V'(\alpha) = \frac{\partial u(\boldsymbol{x}^*, \alpha)}{\partial \alpha} + \sum_{i=1}^{L} \sum_{k=1}^{K} \lambda_k^* \frac{\partial g_k(\boldsymbol{x}^*, \alpha)}{\partial x_i} x_i'(\alpha).$$

Totally differentiating each binding constraint k gives:

$$\sum_{i=1}^{L} \frac{\partial g_k(\boldsymbol{x}^*, \alpha)}{\partial x_i} x_i'(\alpha) + \frac{\partial g_k(\boldsymbol{x}^*, \alpha)}{\partial \alpha} = 0.$$

Multiplying by λ_k^*, summing over k, and changing the order of summation gives:

$$\sum_{i=1}^{L} \sum_{k=1}^{K} \lambda_k^* \frac{\partial g_k(\boldsymbol{x}^*, \alpha)}{\partial x_i} x_i'(\alpha) = -\sum_{k=1}^{K} \lambda_k^* \frac{\partial g_k(\boldsymbol{x}^*, \alpha)}{\partial \alpha}.$$

Hence we have the result. □

4 Duality in Consumer and Firm Theory

4.1 Expenditure Minimization Problem

When economists talk about duality in consumer theory, they normally have in mind the connections between quasiconcave utility maximization in a budget set and expenditure minimization to get to a fixed utility level.[10] Let's start by laying out the expenditure minimization problem for a given utility function $u(\boldsymbol{x})$ on $\mathbb{R}^L_+$:

$$\min_{\{\boldsymbol{x}: u(\boldsymbol{x}) \geq \overline{u}\}} \boldsymbol{p} \cdot \boldsymbol{x} = \sum_{l=1}^{L} p_l x_l.$$

We shall assume throughout that utility functions are strictly monotonic. Let $\boldsymbol{h}(\boldsymbol{p}, \overline{u}) \subset \mathbb{R}^L_+$ denote the set of solutions to this problem. Each selection from this correspondence is called a *compensated or Hicksian demand function* for the problem. The value function $e(\boldsymbol{p}, \overline{u})$ of the problem is called the *expenditure function* of the problem.

Figure 5 for expenditure minimization (or the first order KKT conditions) shows that isoexpenditure sets with prices $\boldsymbol{p}$ and $\lambda \boldsymbol{p}$ are parallel. Hence Hicksian demands are homogenous of degree 0 in $\boldsymbol{p}$ and the expenditure function is homogenous of degree 1 in $\boldsymbol{p}$. You can also show that $e(\boldsymbol{p}, \overline{u})$ is increasing in $\boldsymbol{p}$ and strictly increasing in $\overline{u}$. In fact, we have the following result.

Proposition 4.1. If $e(\boldsymbol{p}, \overline{u})$ is i) homogenous of degree 1 in $\boldsymbol{p}$, ii) concave (and therefore continuous) in p, iii) increasing in $\boldsymbol{p}$, iv) strictly increasing

[10] In mathematical terms, this approach establishes Fenchel duality between each convex upper contour sets and its support function, the expenditure function of the problem. The price vector determines a normal to a hyperplane along which the cost $\boldsymbol{p} \cdot \boldsymbol{x}$ is constant. The expenditure function $e(\boldsymbol{p}, \overline{u})$ determines a constant so that the the hyperplane $\{\boldsymbol{x} \in \mathbb{R}^L_+ | \boldsymbol{p} \cdot \boldsymbol{x} = e(\boldsymbol{p}, \overline{u})\}$ is a supporting hyperplane to the convex set $\{\boldsymbol{x} | u(\boldsymbol{x}) \geq \overline{u}\}$. Since any convex set is determined by its supporting hyperplanes, $e(\boldsymbol{p}, \overline{u})$ contains all the information about the upper contour sets, and therefore the utility function.

in $\overline{u}$. Then it is the value function of an expenditure minimization problem for some quasiconcave utility function u.

The proof is slightly involved so we will not attempt it here.

We saw in the previous section that $e(\boldsymbol{p}, \overline{u})$ is a concave function of $\boldsymbol{p}$. If the Hicksian demand $\boldsymbol{h}(\boldsymbol{p}, \overline{u})$ is a function (i.e. single-valued for all $(\boldsymbol{p}, \overline{u})$, then $e(\boldsymbol{p}, \overline{u})$ is differentiable in $\boldsymbol{p}$ and the envelope theorem gives us the following result:

Proposition 4.2 (Shephard's Lemma)**.** If $e(\boldsymbol{p}.\overline{u})$ is differentiable, then $\frac{\partial e(\boldsymbol{p},\overline{u})}{\partial p_j} = h_j(\boldsymbol{p}, \overline{u})$.

We also note for future use that since $e(\boldsymbol{p}, \overline{u})$ is concave, it has a negative semi-definite Hessian matrix (if it is twice differentiable).

4.1.1 Appendix: Expenditure minimization problem

We cover briefly the mechanics of solving the expenditure minimization problem subject to the constraint of reaching a specified level of utility. All the notation is exactly as in the previous subsection. We assume that the utility function that we have is quasiconcave.

$$\min_{\boldsymbol{x} \in \mathbb{R}^L_+} \boldsymbol{p} \cdot \boldsymbol{x} = \sum_{i=1}^{L} p_i x_i,$$

subject to

$$u(\boldsymbol{x}) \geq \overline{u}.$$

This means that we have a linear and thus quasiconvex objective function for our minimization problem and since the utility function is quasiconcave, the feasible set is convex. Hence we know that KKT necessary conditions are also sufficient. Notice that the feasible set is now not bounded (why?), but the solution exists because we can take any $\boldsymbol{x}^*$ such that $u(x^*) \geq \overline{u}$ and restrict attention to $\boldsymbol{x}$ such that

$$\boldsymbol{p} \cdot \boldsymbol{x} \leq \boldsymbol{p} \cdot \boldsymbol{x}^*,$$

since $\boldsymbol{x}*$ is a feasible solution. But this set is convex and bounded since it is a budget set.

The Lagrangean to the problem is:

$$\mathcal{L}(\boldsymbol{x}, \lambda) = \sum_{i=1}^{L} p_i x_i - \lambda_0(u(\boldsymbol{x}) - \overline{u}) - \sum_{i=1}^{L} \lambda_i x_i.$$

The first-order conditions are:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial x_i} = p_i - \lambda_0 \frac{\partial u}{\partial x_i} - \lambda_i &= 0 \text{ for all } i, \\
\lambda_0[u(x) - \overline{u}] &= 0, \\
\lambda_i x_i &= 0 \text{ for all } i, \\
\overline{u} - u(x) &\leq 0, \\
-x_i &\leq 0 \text{ for all } i, \\
\lambda_i &\geq 0 \;\; i \in \{0, 1, ..., L\}.
\end{aligned}$$

Notice that for interior solutions (where $\lambda_1 = \lambda_2 = ... = \lambda_L = 0$), we get again (after eliminating the multiplier) from the first line of the K-T conditions that

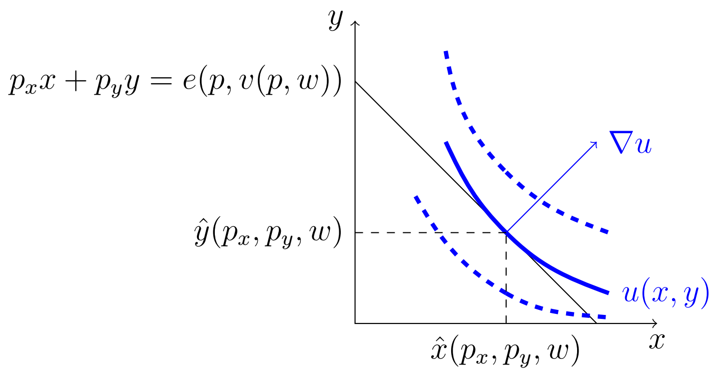
$$\frac{\frac{\partial u(\boldsymbol{x})}{\partial x_i}}{\frac{\partial u(\boldsymbol{x})}{\partial x_k}} = \frac{p_i}{p_k}.$$

We have exactly the same situation as before. Now the ratio of marginal utilities is really the MRT for the problem since it describes the feasible set. The price ratio is now the MRS of this new problem.

4.2 Duality and Slutsky Equation

The Walrasian demand $\boldsymbol{x}(\boldsymbol{p}, w)$ finds the maximal indifference curve in the budget set $B(\boldsymbol{p}, w)$. The value function $v(\boldsymbol{p}, w)$ gives the utility level at this maximal indifference curve. Hicksian demand asks which point on a

Figure 4: UMP for $w = e(\boldsymbol{p}, v(\boldsymbol{p}, w))$



given indifference curve $u(\boldsymbol{x}) = \overline{u}$ is the cheapest at prices $\boldsymbol{p}$. Iso-cost lines (sets of points at equal cost) are uniquely determined by a point and the price vector $\boldsymbol{p}$ (they satisfy $\boldsymbol{p} \cdot \boldsymbol{x} = \overline{c}$ for some cost $\overline{c}$). It should not come as a great surprise that the cheapest way to achieve utility level $v(\boldsymbol{p}, w)$ is by setting $\boldsymbol{h}(\boldsymbol{p}, v(\boldsymbol{p}, w)) = \boldsymbol{x}(\boldsymbol{p}, w)$.

How do we see this? If $\boldsymbol{h}(\boldsymbol{p}, v(\boldsymbol{p}, w)) \not\subset \boldsymbol{x}(\boldsymbol{p}, w)$, then there is some optimal demand $\boldsymbol{h}^*(\boldsymbol{p}, v(\boldsymbol{p}, w)) \in \boldsymbol{h}(\boldsymbol{p}, v(\boldsymbol{p}, w))$ such that $\boldsymbol{h}^*(\boldsymbol{p}, v(\boldsymbol{p}, w)) \notin \boldsymbol{x}(\boldsymbol{p}, w)$. Therefore either $u(\boldsymbol{h}^*(\boldsymbol{p}, v(\boldsymbol{p}, w))) < v(\boldsymbol{p}, w)$ or $\boldsymbol{p} \cdot \boldsymbol{h}^*(\boldsymbol{p}, v(\boldsymbol{p}, w)) > w = \boldsymbol{p} \cdot \boldsymbol{x}(\boldsymbol{p}, w)$, where the last equality is by strict monotonicity of u. In the first case, the target utility is not reached. In the second, the expenditure is not minimized. Hence $\boldsymbol{h}(\boldsymbol{p}, v(\boldsymbol{p}, w)) \subset \boldsymbol{x}(\boldsymbol{p}, w)$

Similar steps allow us show that $\boldsymbol{x}(\boldsymbol{p}, w) \subset \boldsymbol{h}(\boldsymbol{p}, v(\boldsymbol{p}, w))$ and this is left as an exercise.

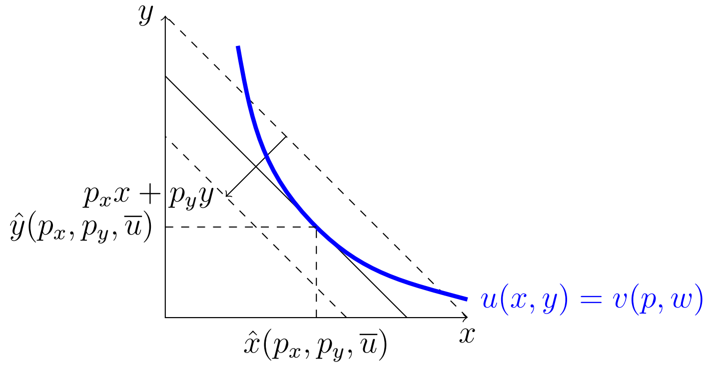
Exercise 4.3. Show that $\boldsymbol{x}(\boldsymbol{p}, w) \subset \boldsymbol{h}(\boldsymbol{p}, v(\boldsymbol{p}, w))$.

I may have been again overly pedantic since in this particular case, one picture is worth a thousand words: see figures 4 and 5.

You can also see that for $\overline{u} = v(\boldsymbol{p}, e(\boldsymbol{p}, \overline{u})$ and $e(\boldsymbol{p}, v(\boldsymbol{p}, w)) = w$ the solutions to expenditure minimization and UMP coincide for all $\boldsymbol{p}$:

$$h_l(\boldsymbol{p}, \overline{u}) = x_l(\boldsymbol{p}, e(\boldsymbol{p}, \overline{u})) \text{ for all } l,$$

Figure 5: Expenditure minimization for $\overline{u} = v(\boldsymbol{p}, e(\boldsymbol{p}, \bar{u}))$



$$h_l(\boldsymbol{p}, v(\boldsymbol{p}, w)) = x_l(\boldsymbol{p}, w) \text{ for all } l.$$

Since these relationships hold for all $\boldsymbol{p}$, the left-hand side and the right-hand side are the same functions when viewed as functions of $\boldsymbol{p}$. We can therefore require all the partial derivatives of the two sides to be equal. Differentiate the first of these identities with respect to p_j to get:

$$\frac{\partial h_l(\boldsymbol{p}, \overline{u})}{\partial p_j} = \frac{\partial x_l(\boldsymbol{p}, w)}{\partial p_j} + \frac{\partial x_l(\boldsymbol{p}, w)}{\partial w} \frac{\partial e(\boldsymbol{p}, \overline{u})}{\partial p_j}.$$

By Shephard's Lemma, $\frac{\partial e(\boldsymbol{p}, \overline{u})}{\partial p_j} = h_j(\boldsymbol{p}, \overline{u})$, so that:

$$\frac{\partial h_l(\boldsymbol{p}, \overline{u})}{\partial p_j} = \frac{\partial x_l(\boldsymbol{p}, w)}{\partial p_j} + \frac{\partial x_l(\boldsymbol{p}, w)}{\partial w} h_j(\boldsymbol{p}, \overline{u}).$$

Using the first relationship above, the right-hand side is

$$\frac{\partial x_l(\boldsymbol{p}, w)}{\partial p_j} + \frac{\partial x_l(\boldsymbol{p}, w)}{\partial w} x_j(\boldsymbol{p}, e(\boldsymbol{p}, \overline{u}))$$

$$= \frac{\partial x_l(\boldsymbol{p}, w)}{\partial p_j} + \frac{\partial x_l(\boldsymbol{p}, w)}{\partial w} x_j(\boldsymbol{p}, e(\boldsymbol{p}, v(\boldsymbol{p}, w))$$

$$= \frac{\partial x_l(\boldsymbol{p}, w)}{\partial p_j} + \frac{\partial x_l(\boldsymbol{p}, w)}{\partial w} x_j(\boldsymbol{p}, w).$$

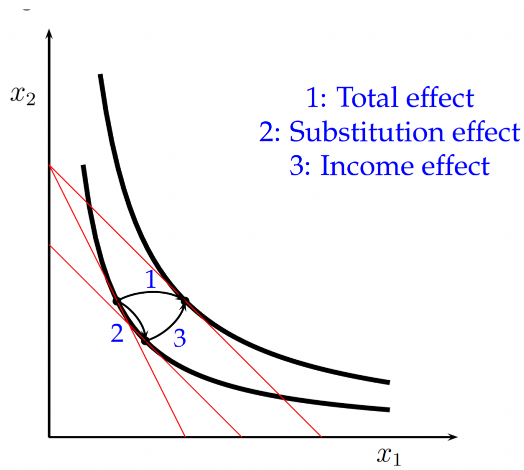


Figure 6: Income and Substitution Effects

This is the famous Slutsky equation for income and substitution effects. The observable change $\frac{\partial x_l(\boldsymbol{p},w)}{\partial p_j}$ in Marshallian demands can be decomposed into a substitution effect, i.e. the change in compensated demand $\frac{\partial h_l(\boldsymbol{p},\overline{u})}{\partial p_j}$ and the observable income effect $\frac{\partial x_l(\boldsymbol{p},w)}{\partial w} x_j(\boldsymbol{p},w)$:

$$\frac{\partial x_l(\boldsymbol{p},w)}{\partial p_j} = \frac{\partial h_l(\boldsymbol{p},\overline{u})}{\partial p_j} - \frac{\partial x_(\boldsymbol{p},w)}{\partial w} x_j(\boldsymbol{p},w).$$

Since we know that the Hessian of $e(\boldsymbol{p},\overline{u})$ is negative semi-definite, we know that its diagonal elements are non-positive. Hence the effect of increasing p_i on x_i is negative whenever the demand for i is increasing in income (we say then that i is a non-inferior good).

4.2.1 Integrability

We have seen that demand functions arising from utility maximization problems satisfy:

1. Homogeneity of degree 0 (budget set does not change if all prices and income multiplied by the same strictly positive number).

2. Walras' law: if the utility function is strictly increasing, then all income is used:

$$\sum_{l=1}^{L} p_l x_l(\boldsymbol{p}, w) = w \text{ for all } \boldsymbol{p} >> 0,\ w > 0.$$

3. The matrix $\boldsymbol{X}$ (called the *Slutsky matrix*) with $(l,k)^{th}$ element $x_{lk} = \frac{\partial x_l(\boldsymbol{p},w)}{\partial p_k} + \frac{\partial x_l(\boldsymbol{p},w)}{\partial w} x_k(\boldsymbol{p}, w)$ is negative semi-definite.

4. From Young's theorem, Slutsky matrix is symmetric.

We could ask conversely, what conditions on a vector valued function $\boldsymbol{x}(\boldsymbol{p}, w)$ guarantee that it is the Marshallian demand for some utility maximization problem. A remarkable (but unfortunately somewhat hard to prove) result states that the above four conditions are sufficient.

Theorem 4.4 (Integrability)**.** If $\boldsymbol{x}(\boldsymbol{p}, w)$ is homogenous of degree 0 in $(\boldsymbol{p}, w)$, satisfies Walras' law and has a symmetric and negative semi-definite Slutsky matrix, then there is a strictly increasing and quasiconcave utility function $u(\boldsymbol{x})$ such that $\boldsymbol{x}(\boldsymbol{p}, w) = \arg\max_{\boldsymbol{x} \in B(\boldsymbol{p}, w)} u(\boldsymbol{x})$.

Now is again a good time to pause and consider what we have achieved. If our standard for coherent choice is that choices maximize a rational preference, we have found a sufficient condition for this. Coherent choice by WARP gave us a negative semi-definite Slutsky matrix. Afriat's theorem demand functions satisfying SARP (i.e. no cycles in indirect revealed preference) gave a condition for a finite set of observations to be consistent with coherent choice.

4.3 Welfare Evaluations and Aggregation

How should we think about the welfare effects of changes in $(\boldsymbol{p}, w)$? As economists, we are often interested in market behavior rather than individual consumers. Do the positive and normative findings on individual buyers generalize to market (aggregate demand) analysis?

Unfortunately we do not have perfect answers to either of the questions. Let's start with individual welfare evaluation. Since utilities are ordinal, we would really want to have the comparisons in monetary terms. One way of doing this is by fixing a price $\overline{\boldsymbol{p}}$ and ask how large expenditure (at those prices) is needed to reach utility level $v(\boldsymbol{p}, w)$. In other words, we would evaluate $e(\overline{\boldsymbol{p}}, v(\boldsymbol{p}, w))$. We would have a fine theory if these measurements were independent of how $\overline{\boldsymbol{p}}$ was chosen, but unfortunately this is not the case.[11]

4.3.1 Individual Welfare Effects

Consider price changes from $\boldsymbol{p}^0$ to $\boldsymbol{p}^1$ holding income w fixed, and denote the attainable utility levels by $u^0 = v(\boldsymbol{p}^0, w)$ and $u^1 = v(\boldsymbol{p}^1, w)$. We could fix the prices for measuring the change either at the old prices $\boldsymbol{p}^0$ or at the new prices $\boldsymbol{p}^1$. The former is called *equivalent variation* or $EV(\boldsymbol{p}^0, \boldsymbol{p}^1, w)$, and the latter is called *compensating variation* or $CV(\boldsymbol{p}^1, \boldsymbol{p}^1, w)$.

$$
\begin{aligned}
EV(\boldsymbol{p}^0, \boldsymbol{p}^1, w) &= e(\boldsymbol{p}^0, u^1) - e(\boldsymbol{p}^0, u^0) = e(\boldsymbol{p}^0, u^1) - w, \\
CV(\boldsymbol{p}^0, \boldsymbol{p}^1, w) &= e(\boldsymbol{p}^1, u^1) - e(\boldsymbol{p}^1, u^0) = w - (\boldsymbol{p}^1, u^0).
\end{aligned}
$$

Since Shephard's Lemma gives: $\frac{\partial e(\boldsymbol{p},u)}{\partial h_l} = h_l(\boldsymbol{p}, u)$ and since $w = e(\boldsymbol{p}^1, u^1) = e(\boldsymbol{p}^0, u^0)$, we can write the above equalities (by the fundamental theorem of calculus) for price changes in p_l as:

$$
\begin{aligned}
EV(\boldsymbol{p}^0, \boldsymbol{p}^1, w) &= \int_{\boldsymbol{p}_1}^{\boldsymbol{p}^0} h_l(\boldsymbol{p}, u^1) dp_l, \\
CV(\boldsymbol{p}^0, \boldsymbol{p}^1, w) &= \int_{\boldsymbol{p}_1}^{\boldsymbol{p}^0} h_l(\boldsymbol{p}, u^0) dp_l.
\end{aligned}
$$

For normal goods, $h_l(\boldsymbol{p}, u^0) < h_l(\boldsymbol{p}, u^1)$ if $u^0 < u^1$. Hence we see that for normal goods, the welfare effect for a decrease in p_l satisfies:

$$
CV(\boldsymbol{p}^0, \boldsymbol{p}^1, w) < EV(\boldsymbol{p}^0, \boldsymbol{p}^1, w).
$$

[11] This is also the reason why it is not possible to construct perfect index numbers for welfare measurement.

If there are no income effects (i.e. demand for good l is independent of income), then the two measures coincide.

A third alternative (and one that is often used in practice since the Walrasian demands are easier to estimate than Hicksian demands) is *consumer surplus* denoted by $CS(\boldsymbol{p}^0.\boldsymbol{p}^1, w)$, and defined (for changes in p_l) by:

$$CS(\boldsymbol{p}^0, \boldsymbol{p}^1, w) = \int_{\boldsymbol{p}_1}^{\boldsymbol{p}^0} x_l(\boldsymbol{p}, w) dp_l.$$

For $u^1 > u^0$, we have for normal goods:

$$h_l(\boldsymbol{p}, u^0) < x_l(\boldsymbol{p}, w) < h_l(\boldsymbol{p}, u^1),$$

(with reverse inequalities for inferior goods), and we get:

$$CV(\boldsymbol{p}^0, \boldsymbol{p}^1, w) < CS(\boldsymbol{p}^0, \boldsymbol{p}^1, w) < EV(\boldsymbol{p}^0, \boldsymbol{p}^1, w),$$

and all three measures coincide if there are no income effects.

Again, it is a good idea to look at a picture. For normal goods, Slutsky equation tells us that Hicksian demands are steeper than Walrasian ones.

Exercise 4.5. Locate the two welfare measures in Figure 7. Where is the Walrasian demand located in this Figure?

4.4 Cost minimization problem for a firm

A firm chooses its inputs k, l to minimize the cost of reaching a production target of $\overline{q}$ at given input prices r, w. The production function is assumed to be a strictly increasing and quasiconcave function $f(k, l)$.

$$\min_{(k,l)\in\mathbb{R}^2_+} rk + wl$$

subject to

$$f(k, l) \geq \overline{q}.$$

Notice that this is the same mathematical problem as in expenditure minimization. Only the names of variables have changed. The solution to the problem is therefore also identical and we do not repeat it here.

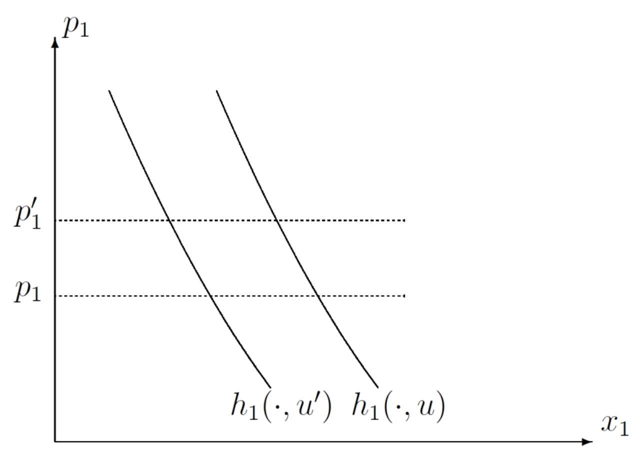


Figure 7: Hicksian Demands at Different Utility levels

4.5 Computed Example: All the Way to Slutsky

Start with a utility maximization problem with Cobb-Douglas preferences and strictly positive prices p_x, p_y for the two goods x, y, i.e.:

$$\max_{x,y} x^{\alpha} y^{1-\alpha}$$

subject to:

$$p_x x + p_y y \leq w, x \geq 0, y \geq 0.$$

Form the Lagrangean:

$$\mathcal{L}(x, y, \lambda_1, \lambda_2, \lambda_3) = x^{\alpha} y^{1-\alpha} - \lambda_1(p_x x + p_y y - w) + \lambda_2 x + \lambda_3 y.$$

We have argued in previous lectures that since the utility function is strictly increasing, the budget constraint will bind and the non-negativity constraints do not bind at the optimum. We have also derived the solution to be:

$$x(p_x, p_y, w) = \frac{\alpha w}{p_x}, y(p_x, p_y, w) = \frac{(1-\alpha)w}{p_y}.$$

By substituting these optimal solutions to the objective function, we get the indirect utility function:

$$v(p_x, p_y, w) = x(p_x, p_y, w)^{\alpha} y(p_x, p_y, w)^{(1-\alpha)} = w(\frac{\alpha}{p_x})^{\alpha}(\frac{(1-\alpha)}{p_y})^{(1-\alpha)}.$$

You should check that this indirect utility function is homogenous of degree 0 in (p_x, p_y, w), i.e. that multiplying both prices and the income w by the same positive number λ leaves that value of the indirect utility function unchanged. You should also check that Roy's identity holds, i.e. that you get the demand function from:

$$x(p_x, p_y, w) = -\frac{\frac{\partial v(p_x,p_y,w)}{\partial p_x}}{\frac{\partial v(p_x,p_y,w)}{\partial w}}.$$

Consider next the expenditure minimization problem for the same preferences:

$$\min_{x,y} p_x x + p_y y$$

subject to:

$$x^{\alpha} y^{1-\alpha} \geq \bar{u}, x \geq 0, y \geq 0.$$

Form the Lagrangean (to get the signs of the multipliers correct, you may remember that minimizing $f(x)$ has the same solution as maximizing $-f(x)$
:

$$\mathcal{L}(x, y, \lambda_i, \lambda_2, \lambda_3) = -p_x x - p_y y + \lambda_1(x^{\alpha} y^{1-\alpha} - \bar{u}) + \lambda_2 x + \lambda_3 y.$$

We see immediately that if $\bar{u} = 0$, then the optimal solution is $x = y = 0$. For $\bar{u} > 0$, the only feasible consumptions are interior and hence $\lambda_2 = \lambda_3 = 0$. The utility constraint must be binding since otherwise it would be possible to lower the consumption of one of the goods leading to a smaller expenditure without violating any of the constraints.

The first order conditions for this minimization problem are:

$$p_x - \lambda_1 \alpha \frac{x^{\alpha} y^{1-\alpha}}{x} = 0,$$

$$p_y - \lambda_1(1-\alpha)\frac{x^\alpha y^{1-\alpha}}{y} = 0,$$
$$x^\alpha y^{1-\alpha} - \bar{u} = 0.$$

Solving from the first two equations, we get:

$$y = \frac{p_x}{p_y}\frac{(1-\alpha)}{\alpha}x.$$

Substituting to the third first-order condition gives:

$$x(p_x, p_y, \bar{u}) = \bar{u}(\frac{p_x}{p_y}\frac{(1-\alpha)}{\alpha})^{\alpha-1},$$

and similarly:

$$y(p_x, p_y, \bar{u}) = \bar{u}(\frac{p_y}{p_x}\frac{\alpha}{(1-\alpha)})^{-\alpha}.$$

These are called the Hicksian or compensated demands for x and y. The value function, i.e. the expenditure function is then:

$$e(p_x, p_y, \bar{u}) = \bar{u}(p_x(\frac{p_x}{p_y}\frac{(1-\alpha)}{\alpha})^{\alpha-1} + p_y(\frac{p_y}{p_x}\frac{\alpha}{(1-\alpha)})^{-\alpha}).$$

Is the expenditure function homogenous? Of what degree? Can you see that by taking the partial derivative of this expenditure function, you get back the compensated demand for x?

Finally, you can verify the Slutsky equation. For example, the partial derivative of the compensated demand for x with respect to own price:

$$\frac{\partial x(p_x, p_y, \bar{u})}{\partial p_x} = \bar{u}(\alpha-1)\frac{1}{p_x}(\frac{p_x}{p_y}\frac{(1-\alpha)}{\alpha})^{\alpha-1}$$

is equal to

$$\frac{\partial x(p_x, p_y, w)}{\partial p_x} + x(p_x, p_y, w)\frac{\partial x(p_x, p_y, w)}{\partial w} = -\frac{\alpha w}{p_x^2} + \frac{\alpha^2 w}{p_x^2},$$

when evaluated at $\bar{u} = v(p_x, p_y, w) = w(\frac{\alpha}{p_x})^\alpha(\frac{(1-\alpha)}{p_y})^{(1-\alpha)}$.

4.6 Aggregation

Intermediate microeconomics is centered around the questions of market demand and market supply. What happens when we look at aggregate demand, i.e. when we sum together individual demands? Will the aggregate demand satisfy the properties that we derived for the individual demands? If all consumers face the same prices, does the distribution of wealth amongst the consumers affect the aggregate demand? Can we have a welfare interpretation for the market demand, i.e. can market demand be assumed to arise from a representative consumer?

We will not have time for a full discussion of these issues in these notes, but I will comment briefly on each question.

If all consumers $i \in \{1, ..., I\}$ in the economy have Walrasian demand functions $\boldsymbol{x}_i(\boldsymbol{p}, w_i)$ derived, does the aggregate demand $\boldsymbol{x}(\boldsymbol{p}, \boldsymbol{w}) = \sum_{i=1}^{I} \boldsymbol{x}_i(\boldsymbol{p}, w_i)$, where $\boldsymbol{w} = \sum_{i=1}^{I} w_1, ..., w_I$ satisfy the necessary conditions for rational demand functions? Unfortunately this fails badly. We cannot even get compensated law of demand for the sum of individual demands that satisfy it. Here is a simple example showing this

Example 4.6. Assume two consumers with equal income $w = 4.4$. At prices $\boldsymbol{p} = (2, 1)$ consumer 1 demands $\boldsymbol{x}^1(2, 1; 4.4) = (1.4, 1.6)$ and at prices $\boldsymbol{p} = (1, 2)$ she demands $\boldsymbol{x}^1(1, 2; 4.4) = (0, 2.2)$. For consumer 2, the demands are $\boldsymbol{x}^2 = (2, 1; 4.4) = (2.2, 0)$ and $\boldsymbol{x}^2(1, 2; 4.4) = (1.6, 1.4)$ respectively. The aggregate demands $\boldsymbol{x}(2, 1, 8.8) = (3.6, 1.6)$ and $\boldsymbol{x}(1, 2, 8.8) = (1.6, 3.6)$ do not satisfy WARP (or if you will, the average demand is not consistent with WARP).

On the other hand, if all buyers have Walrasian demand functions satisfying the individual law of demand $\frac{\partial \boldsymbol{x}_{il}(\boldsymbol{p}, w_i)}{\partial p_l} < 0$, then aggregate law of demand holds also: $\frac{\partial \boldsymbol{x}_l(\boldsymbol{p}, \boldsymbol{w})}{\partial p_l} < 0$. This follows immediately from the fact that the derivative of a sum is the sum of derivatives.

For the second question regarding the sensitivity to changes in income distribution, the only instance where aggregation works nicely is if all consumers have expenditure functions that take the Gorman polar form:

$$e_i(\boldsymbol{p}, u_i) = a_i(\boldsymbol{p}) + b(\boldsymbol{p})u_i.$$

This can be interpreted as saying that the consumers have parallel linear (or affine) *Engel curves*, i.e. income expansion paths. To see this, use $e_i(\boldsymbol{p}, v(\boldsymbol{p}, w_i)) = w_i$ and $u_i = v_i(\boldsymbol{p}, w_i)$ to invert:

$$v_i(\boldsymbol{p}, w_i) = \frac{w_i - a_i(\boldsymbol{p})}{b(\boldsymbol{p})}.$$

Then use Roy's identity to get individual demand for good l:

$$x_{il} = \frac{\partial a_i(\boldsymbol{p})}{\partial p_l} + (w_i - a_i(\boldsymbol{p}))\frac{\frac{\partial b(\boldsymbol{p})}{\partial p_l}}{b(\boldsymbol{p})},$$

which is clearly linear in w_i. Since $\frac{\frac{\partial b(\boldsymbol{p})}{\partial p_l}}{b(\boldsymbol{p})}$ is the same for all consumers, only aggregate income, not its distribution, matters for aggregate demand.

When using this result, you should be very careful with the sign on the derived demands. In most cases, only positive demands make sense. Incorporating this requirement may result in a piece-wise linear Engel-curve, and then the aggregation result does not hold.

Particular cases of Gorman polar form include the setting where consumers have identical homothetic utility functions so that $v_i(\boldsymbol{p}, w_i) = v(\boldsymbol{p})w_i$, quasilinear utilities and the Stone-Geary utility function that we already saw in 3.3.3.

Finally for the last point, the results are quite disappointing. Apart from the case of identical consumers, or consumers with expenditure functions in the Gorman polar form, very few cases admit a positive or a normative interpretation of the market demand as the demand of a meaningfully defined representative consumer. See MWG chapter 4 for additional material.

4.7 Appendix: Other Dualities

Another duality relationship in classical consumer theory is the duality between utility maximization and indirect utility minimization. This is closer to the duality that you may have seen in Linear Programming. Since $B(\tilde{\boldsymbol{p}}, w) = B(\frac{\tilde{\boldsymbol{p}}}{w}, 1)$, we can let $w = 1$ for this discussion and just talk about $\boldsymbol{p} = \frac{\tilde{\boldsymbol{p}}}{w}$ so that $v(\boldsymbol{p}) = \max_{\boldsymbol{x} \in B(\boldsymbol{p})} u(\boldsymbol{x})$. The only point where you need to be careful is with the allowed forms of utility functions. If the utility function is quasiconcave, you can verify the following by drawing a picture (or by considering the KKT conditions for the minimization problem).

$$v(\boldsymbol{p}^*(\boldsymbol{x}^*)) = u(\boldsymbol{x}^*(\boldsymbol{p}^*),$$

where $\boldsymbol{p}^*$ solves

$$\min_{\{\boldsymbol{p}:\boldsymbol{p}\cdot\boldsymbol{x}^*=1\}} v(\boldsymbol{p}),$$

and $\boldsymbol{x}^*$ solves

$$\max_{\{\boldsymbol{x}:\boldsymbol{p}^*\cdot\boldsymbol{x}=1\}} u(\boldsymbol{x}).$$

Based on this duality, you will believe that an indirect utility function encodes all information present in a quasiconcave utility function, and you can find the Walrasian demands by differentiation (Roy's identity).

Let us summarize: If a function $v(\boldsymbol{p}, w)$ is an indirect utility function of a utility maximization problem for some strictly monotone quasiconcave and continuous utility function u, then it is: i) Homogenous of degree 0 in $(\boldsymbol{p}, w)$, ii) Strictly decreasing in p_l for all l when $w > 0$, iii) Strictly increasing in w, iv) quasiconvex in $(\boldsymbol{p}, w)$, v) Continuous.

Conversely any function $v(\boldsymbol{p}, w)$ satisfying these properties is an indirect utility function for some strictly monotonic continuous and quasiconcave utility function.

4.8 Duality in Firm Theory

There are two useful dualities in the theory of the firm. We start with the most fundamental representation of the profit maximization problem.

$$\max_{\boldsymbol{y}\in Y} \boldsymbol{p}\cdot\boldsymbol{y}.$$

We denote the solution to this problem by the supply correspondence $\boldsymbol{y}(\boldsymbol{p})$, and the value function of the problem ise denoted by $\pi(\boldsymbol{p})$ and called the *profit function*. The duality is now between the production set Y and its support function, the profit function. If Y is convex, then we have the duality:

$$Y = \{\boldsymbol{y}|\boldsymbol{p}\cdot\boldsymbol{y} \leq \pi(\boldsymbol{p}) \text{ for all } \boldsymbol{p}\}.$$

We have again by envelope theorem the following proposition.

Proposition 4.7 (Hotelling's Lemma)**.** If $\pi(\boldsymbol{p})$ is differentiable, then $\frac{\partial\pi(\boldsymbol{p})}{\partial p_l} = y_l(\boldsymbol{p})$.

The proof is left as an exercise. The other important duality is for the case where the firm has a single output y produced from input vector $\boldsymbol{z}$ according to the quasiconcave production function $y = f(\boldsymbol{z})$. Let p be the output price and $\boldsymbol{w}$ the vector of input prices. Then we can look at the cost minimizing problem subject to producing at least q units.

$$\min_{\{\boldsymbol{z}:f(\boldsymbol{z})\geq q\}} \boldsymbol{w}\cdot\boldsymbol{z}.$$

Apart from changing the names of variables, you should recognize this as the same mathematical problem as the expenditure minimization problem that we had before. $\boldsymbol{z}(\boldsymbol{w};q)$ is known as the conditional input demand and the value function $c(\boldsymbol{w},q) = \min_{\{\boldsymbol{z}:f(\boldsymbol{z})\geq q\}} \boldsymbol{w}\cdot\boldsymbol{z}$ is called the cost function. One key difference is that f is a function determined by the production technology and therefore it is cardinal. As a result, if we look at increasing transformations of f, we have different technologies.

Notice that in this course, we have been really interested in how c behaves as a function of $\boldsymbol{w}$, i.e. the comparative statics representing changes between technologies as a function of input prices. In intermediate microeconomics courses, the interest is mostly on how c depends on q (the prices are kept fixed throughout the analysis).

Of course at the end of the day, the firm decides optimal level of production:

$$\max_{q \in \mathbb{R}_+} pq - c(\boldsymbol{w}; q).$$

At this step, we get back the familiar intermediate microeconomics first-order condition for profit maximization:

$$p = \frac{\partial c(\boldsymbol{w}; q)}{\partial q}.$$

5 Choice under Uncertainty and Risk

Uncertainty plays a major role in economic decision making. Finance is all about handling uncertainties. Real investments, career choices, and retirement savings decisions all involve a major element of uncertainty. Up to now, we have talked about choices from sets of certain alternatives. In this section, we want to expand the scope of our theory to cover uncertain outcomes as well.

You will see a modeling strategy that may seem a bit confusing at first sight. We expand the set over which choices are made and on which preferences are formulated. We call this set the set of *lotteries*, and will then impose assumptions beyond the standard transitivity, continuity and monotonicity assumptions to develop a normatively attractive and analytically tractable model of choice under uncertainty. But before starting this, we discuss different interpretations we could have for uncertainty.

5.1 Uncertainty and Risk

The mathematical model for uncertainty is developed in probability theory. We have a sample space Ω and a collection $\mathcal{E}$ of events, i.e. sets $E \subset \Omega$ defined on it. A probability (measure) attaches to each $E \in \mathcal{E}$ a number $0 \leq p(E) \leq 1$. The empty set $\emptyset$ and the whole sample space Ω are events, and $p(\emptyset) = 0, p(\Omega) = 1$. $\mathcal{E}$ is closed under unions and intersections, i.e. if $A, B \in \mathcal{E}$, then $A \cup B \in \mathcal{E}$, and $A \cap B \in \mathcal{E}$. [12]

Probability is additive in the sense that if $A, B \in \mathcal{E}$ and $A \cap B = \emptyset$, then $p(A \cup B) = p(A) + p(B)$.[13] At this point, it is a good idea to review the basic facts of probability theory including: random variables, expectations, independence, conditional probability, Bayes' rule.

[12] For the purposes of this subsection, you can safely assume that Ω is finite and that $\mathcal{E}$ is the collection of all of its subsets.

[13] When Ω is infinite, then a stronger form of additivity is needed to develop the theory properly.

The easiest way of thinking about uncertainty for decision making is the model of objectively given probabilities. Objective uncertainty is often referred to as risk, while uncertainty is a term left for subjective probabilities (see below). In the model of risk, a decision maker considers each $\omega \in \Omega$ as a priori possible and assigns an exogenously given probability $p(\omega)$ to each sample point (and by finiteness, and additivity, these probabilities determine $p(E)$ for all events, i.e. all subsets $E \in \Omega$). Our first model develops optimal choice theory for risky lotteries, where we attach to each ω a consequence $x(\omega)$, and derive a an expected utility representation for preferences over probabilistic consequences. An expected utility representation constructs a utility function on consequences with the particular property that any distribution (i.e. lottery) on consequences is evaluated as a probability weighted average of the utilities on consequences.

It is clear that well defined exogenous probabilities do not exist for all uncertain prospects. Even experts disagree on the distribution of Tesla's profit for the next quarter or the probability of Trump being elected in November 2024 (this is written in August 2024). If we want to encompass such subjective uncertainties, we need an altogether different setting.

Following de Finetti and Savage, we can take a revealed preference approach to choice under uncertainty. The key object for this analysis is the set of *acts*, i.e. functions $f \in F$ mapping the set of *states* $s \in S$ to *consequences* $x \in X$ so that $f(s)$ is the consequence under act f in state s. Savage imposed a set of axioms on the preferences on F that result in an expected utility representation that includes a subjective probability distribution on the states $s \in S$ and a utility function on the consequences. The theory of subjective expected utility is one of the gems of mathematical economics, but unfortunately we do not have the time (or the pre-requisites) for a full treatment.[14]. For our purposes, it is good to note that both the objective and subjective approach leads to a mathematical theory where uncertain

[14] *Utility Theory for Decision Making* by Peter Fsihburn (1970) is an authoritative source for this.

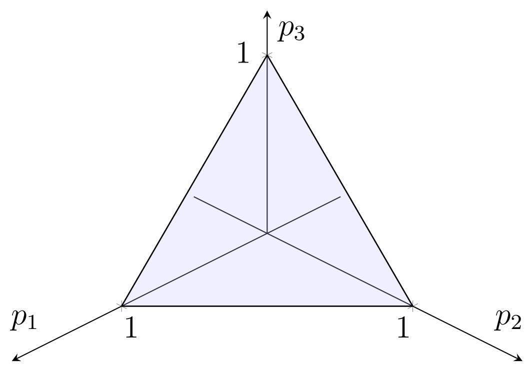


Figure 8: Probabilities in $\mathbb{R}^3$

consequences are evaluated using an expected utility formula.

5.2 Expected Utility Theorem

A *simple lottery* on the (finite) set of consequences $X = \{x_1, ..., x_K\}$ is a non-negative vector p in $\mathbb{R}^K_+$ whose coordinates sum to 1.[15] The set of lotteries is denoted by:

$$\mathcal{L} = \{p = (p_1, ..., p_K) | \sum_{i=1}^{K} p_i = 1,\ p_i \geq 0 \text{ for all } i.\}$$

We also use the notation $\Delta(X)$ to denote the set of probabilities over a set X so that $\mathcal{L} = \Delta(X)$. We denote the degenerate (i.e. certain) lotteries by $\delta_k = (0, ..., 0, 1, 0, ...0)$, where the only non-zero term is in the $k^t h$ coordinate. The degenerate lottery δ_k delivers consequence x_k for sure. For three consequences, you can visualize the set of lotteries as the 2-dimensional simplex in $\mathbb{R}^3$ as in Figure 8.

Let $\succeq$ be a continuous rational preference relation on $\mathcal{L}$. Since X is finite, we can find degenerate lotteries δ_B such that $\delta_B \succeq \delta_k$ for all k and δ_W such that $\delta_k \succeq \delta_W$ for all k.

[15]Please note that I use p without subscripts for the vector of probabilities. Do not confuse this with the price vector in previous sections.

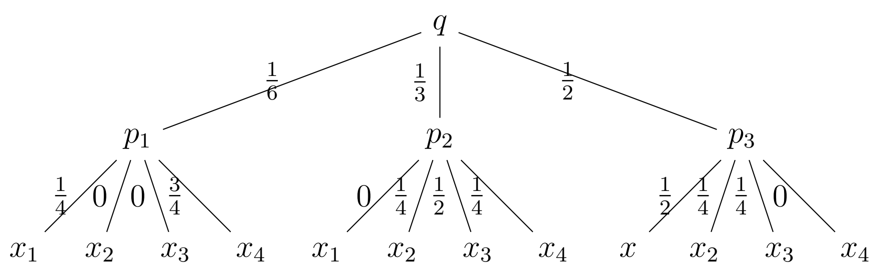


Figure 9: Compound Lottery $q \circ p$

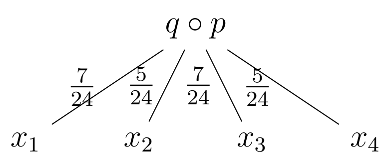


Figure 10: Simple Lottery Induced by $q \circ p$

The idea here is to use the best and worst lotteries to calibrate appropriate utilities for all lotteries. Before that, we must talk about compound lotteries. By this, we simply mean lotteries over lotteries, i.e. probability distributions over different lotteries. A compound lottery $q \circ p \in \Delta(\Delta(X))$ is a (random variable with finite support in $\Delta(X)$). If we have a compound lottery with first stage lottery $q = (q_1, ..., q_L)$ and each of the second stage lotteries $p_l \in \Delta(X)$.

Perhaps it is easiest to explain through an example. A compound lottery $q \circ p$ with $q = (q_1, q_2, q_3) = (\frac{1}{6}, \frac{1}{3}, \frac{1}{2})$ and $p_1 = (\frac{1}{4}, 0, 0, \frac{3}{4}), p_2 = (0, \frac{1}{4}, \frac{1}{2}, \frac{1}{4}), p_3 = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4}, 0)$on $X = \{x_1, x_2, x_3, x_4\}$ gives the final consequences (also called prizes) x_k with probabilities $p(x_k) = \sum_{l=1} q_l p_{lk}$. So for example prize x_3 is obtained with probability $p(x_3) = \frac{1}{6} \times 0 + \frac{1}{3} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{4} = \frac{7}{24}$.

Or perhaps a picture may help: see figures 9 and 10.

5.2.1 The Independence Axiom

The classic result in choice under uncertainty, the *Expected Utility representation* for preferences over lotteries hinges on the following axiom. We write $\alpha p \oplus (1-\alpha)p'$ for the simple lottery on X derived from the compound lottery that gives lottery p with probability α and lottery p' with the complementary probability $(1-\alpha)$.

Assumption 5.1 (Independence)**.**
A rational preference on $\mathcal{L}$ satisfies the *independence axiom* if for all $p, q, r \in \mathcal{L}$, and for all $\alpha \in [0,1]$, we have

$$\alpha p \oplus (1-\alpha)r \succeq \alpha q \oplus (1-\alpha)r \iff p \succeq q.$$

The idea here is that with probability $(1-\alpha)$, the compound lottery yields the same simple lottery r in both cases. Therefore only the part that is different should matter, i.e. the comparison for the two compound lotteries should reduce to the comparison of p versus q. Note that this assumption would be very restrictive (and sometimes quite strange) in our previous sections on choice. It is hard to see why such independence should hold if α is regarded as mixing in terms of quantities of bundles p, q with another bundle r.

The following observation on preferences satisfying the independence axiom is useful in what follows.

Lemma 5.2. Suppose that $\succeq$ satisfies the independence axiom and $p \succeq q$. Then $\alpha p \oplus (1-\alpha)q \succeq \beta p \oplus (1-\beta)q \iff \alpha \geq \beta$.

Exercise 5.3. Prove Lemma 5.2.

Theorem 5.4. A rational and continuous preference relation $\succeq$ on $\mathcal{L}$ satisfies the independence axiom if and only if there exists a utility function $u : X \to \mathbb{R}$ such that

$$p \succeq q \iff U(p) = \sum_{k=1}^{K} p_k u(x_k) \geq U(q) = \sum_{k=1}^{K} q_k u(x_k).$$

Proof. Since X is finite, there exist some best and worst certain lotteries δ_B and δ_W. If $\delta_B \sim \delta_W$, all lotteries are equally good and we set $u(\delta_k) = 1$ for all k.

Suppose then that $\delta_B \succ \delta_W$ and set $u(\delta_W) = 0$ and $u(\delta_B) = 1$. For each δ_k, find α_k such that $\alpha_k \delta_B + (1 - \alpha_k)\delta_W \sim \delta_k$. We know that $U_k = \{\alpha \in [0,1] | \alpha\delta_B + (1-\alpha)\delta_W \succeq \delta_k\}$ and $W_k = \{\alpha \in [0,1] | \delta_k \succeq \alpha\delta_B + (1-\alpha)\delta_W\}$ are non-empty (by definition of δ_B and δ_W, $1 \in U_k$ and $0 \in W_k$) and closed (since $\succeq$ is continuous). By independence axiom, U_k and W_k are intervals. Since $[0,1]$ is a closed connected interval, so is $I_k = U_k \cap W_k$. Independence axiom guarantees again that I_k is a singleton that we call α_k.

We claim that for $u(x_k) = \alpha_k$, the function$U(p) = \sum_{k=1}^{K} p_k u(x_k)$ represents $\succeq$. Since each $\delta_k \sim \alpha_k\delta_B + (1-\alpha_k)\delta_W$, we can view p as a compound lottery that reduces to a simple lottery that gives the prize δ_B with probability $\sum_{k=1}^{K} p_k\alpha_k$ and δ_W with probability $(1 - \sum_{k=1}^{K} p_k\alpha_k)$. We have

$$\sum_{k=1}^{K} p_k\alpha_k \, \delta_B \oplus (1 - \sum_{k=1}^{K} p_k\alpha_k) \, \delta_W \sim p$$

$$\succeq q \sim \sum_{k=1}^{K} q_k\alpha_k \, \delta_B \oplus (1 - \sum_{k=1}^{K} q_k\alpha_k) \, \delta_W.$$

By Lemma 5.2, we have $p \succeq q \iff \sum_{k=1}^{K} p_k\alpha_k\delta_B \geq \sum_{k=1}^{K} q_k\alpha_k\delta_B$. □

Here is a bit of terminology. The utility function U on $\mathcal{L}$ is called the von Neumann-Morgenstern utility function. The utility function u on X is called the Bernoulli utility function. The content of the expected utility theorem is that if the independence axiom holds, then the preference over lotteries can be deduced from the preference over consequences by just computing the expectation of the utilities over consequences. This is a huge simplification of the problem.

Notice that the preference U is linear in probabilities $(p_1, ..., p_K)$. This implies that the indifference curves (for lotteries) are parallel hyperplanes in the simplex $\Delta(X)$. Perhaps the easiest case to visualize is with three

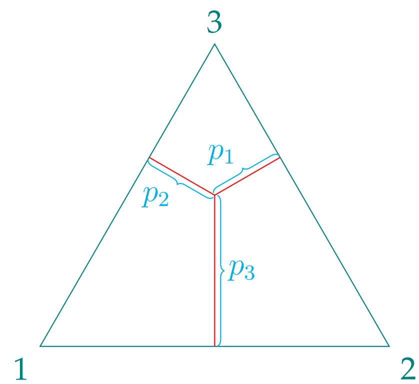


Figure 11: Probability Triangle

consequences so that $p = (p_1, p_2, 1 - p_1 - p_2)$. We can visualize this in an equilateral triangle of height 1, where we identify the corners of the triangle as the degenerate lotteries δ_k for $k = 1, 2, 3$.

One key difference to choice under certainty is that the utility representation is no longer purely ordinal. The following example demonstrates this.

Example 5.5. Suppose $X = \{1, 2, 3\}$ and $u(x) = x$. Then $\delta_2 \sim \frac{1}{2}\delta_1 \oplus \frac{1}{2}\delta_3$. Consider then a strictly increasing transformation $v(u(x)) = (u(x))^2 = x^2$. Then $\frac{1}{2}v(1) + \frac{1}{2}v(3) = \frac{1}{2} \times 1 + \frac{1}{2} \times 9 > v(2) = 4$ showing that u and v represent different preferences.

It is clear that expected utility representations are not unique. If $b > 0$, then $U(p) \geq U(q)$ if and only if $a + bU(p) > a + bU(q)$. Hence U and $V = a + bU$ represent the same preferences. The following proposition shows that expected utility representations are unique only up to such transformations.

Proposition 5.6. Bernoulli utility functions u and v represent the same preferences if and only if for all k, $v(x_k) = a + bu(x_k)$ for some $b > 0$.

Proof. If u and v are two Bernoulli utility functions, let U and V be the corresponding von Neumann-Morgenstern utility functions. The claim is proved if we prove it for U and V.

Let $\alpha(p)$ solve

$$U(p) = \alpha(p)U(\delta_B) + (1 - \alpha(p))U(\delta_W),$$

so that

$$\alpha(p) = \frac{U(p) - U(\delta_W)}{U(\delta_B) - U(\delta_W)}.$$

Since V is also a representation of the same preferences, we have:

$$V(p) = \alpha(p)V(\delta_B) + (1 - \alpha(p))V(\delta_W).$$

Plugging in the value of $\alpha(p)$ and rearranging, we get:

$$V(p) = bU(p) + a,$$

where

$$b = \frac{V(\delta_B) - V(\delta_W)}{U(\delta_B) - U(\delta_W)},$$

and

$$a = V(\delta_W) - U(\delta_W)\frac{V(\delta_B) - V(\delta_W)}{U(\delta_B) - U(\delta_W)}.$$

□

Exercise 5.7. Suppose a decision maker can invest her initial wealth of 1 euro in two different equities $i \in \{1, 2\}$. Each equity sells at 1 euro and equity i delivers x_i euros with probability p_i and 0 with the complementary probability $(1 - p_i)$. Describe the possible portfolios for different investment strategies (you do not have to invest all your wealth in the assets but just keep all of it or part of it uninvested) and compute the probability distributions on final wealth assuming that the assets are statistically independent.

Another nice way of writing the expected utility formula involves a change of variable. If X is a random variable on the real line with distribution function F, then the standard way of writing the expected utility formula is:

$$\mathbb{E}U = \int_{\underline{x}}^{\overline{x}} u(x)dF(x).$$

If we change variables to $s = F(x)$, we can write this formula as:

$$\mathbb{E}U = \int_0^1 u(F^{-1}(s))ds.$$

The variable s has the natural interpretation as the quantile of the distribution F and we see that the expected utility formula gives an expectation of u evaluated at a uniformly distributed random quantile. We may write in general $f_X(s) = F^{-1}(s)$ for the quantile function of X.

5.2.2 Some Extensions

We have assumed thus far that X is finite. For countably infinite X we need to worry about the convergence of sums of type $\sum_{k=1}^{\infty} p_k u(x_k)$. St. Petersburg's paradox is a prime example of this. We could ask the following: how much would you pay for a gamble that pays off 2^k euros with probability $\frac{1}{2^k}$. To get a better sense of the gamble, you can think of flipping a coin to determine the winnings. Your winnings are 2 to the power of the number of consecutive heads that you get in a sequence of coin flips. The expected value of this gamble is infinite. Because of this, Jacob Bernoulli (living in St. Petersburg) suggested that we should consider the expectation of a utility function on the winnings $u(x_k)$ rather than the expected value of the winnings. If u is bounded, all sums of the type above do converge.

The procedure for calibrating the utility function has to be modified slightly since now there does not necessarily exist best and worst prizes. Fortunately enough, this is not too hard. Pick two prizes x_0 and x_1 such that $x_1 \succ x_0$. Put $u(x_1) = 1,\ u(x_0) = 0$. For any other prize x_k find α_k as before if $x_1 \succeq x_k \succeq x_0$. If $x_k \succ x_1$, define α_k by $\delta_1 \sim \alpha_k \delta_k \oplus (1-\alpha_k)\delta_0$, and set $u(x_k) = \frac{1}{\alpha_k}$. If $x_0 \succ x_k$, define α_k by $\delta_0 \sim \alpha_k \delta_k \oplus (1-\alpha_k)\delta_1$, and set $u(x_k) = -\frac{1-\alpha_k}{\alpha_k}$.

For an uncountably infinite X, lotteries must be measurable random variables and the convergence of the integrals needed for computing expectations becomes more tricky. Again, problems can be avoided if the lotteries are on a bounded support and the Bernoulli utility function is continuous on the support. Many convenient settings fail these sufficient conditions (e.g. normally distributed lotteries and exponential utilities), but still yield well-defined expected utilities.

In these notes, we will have only parametric examples where you can use the familiar computations from elementary (i.e. without measure theory) probability. The only reason for considering such extensions is that for random variables with continuous densities on the real line, computations become a lot easier (e.g. we can use calculus to characterize first-order conditions).

Exercise 5.8. Consider normally distributed lotteries, $\tilde{x} \sim \mathcal{N}(\mu, \sigma^2)$ and an exponential utility function on consequences $u(x) = -\frac{1}{\gamma}e^{-\gamma x}$. Compute the expected utility for such lotteries.

If you want a more advanced treatment of the expected utility theorem, the notes by Strzalecki (chapter 5) give a very nice treatment. The idea is to prove first the linear form of the representations, i.e. $U(\lambda p \oplus (1-\lambda)r) = \lambda U(p) + (1-\lambda)U(r)$. The representation in this form extends nicely to much more general spaces of lotteries.

5.2.3 Allais' Paradox

Experimental evidence suggests that expected utility theory fails in some treatments. Perhaps the most famous (and oldest) example of this is known as Allais' Paradox. Here is one version of it.

A decision maker chooses between two lotteries p, q with prizes in $x \in \{2500, 2400, 0\}$. You should interpret these as monetary rewards. Lottery p has probabilities $p = (.33, .66, .01)$ on the three prizes while lottery $q = (0, 1, 0)$. Which would you choose? Why? The decision maker also chooses between r, s on the same prizes with probabilities $r = (.33, 0, .67)$ and $s = (0, .34, .66)$. What would you choose? Why?

Many decision makers say they would choose q over p to avoid the possibility of the zero prize. At the same time, they would choose r over s since $.33$ is not very different from $.34$ and the prize in the first case is larger than in the second.[16]

Is this consistent with expected utility maximization? If you prefer q over p, then $u(2400) > .33u(2500) + .66u(2400) + .01u(0)$. If you prefer r over s, then $.33u(2500) + .67u(0) > .34u(2400) + .66u(0)$. These are clearly inconsistent.

Let's see if we can find a violation of the independence axiom. The first choice is between $.66(0, 1, 0) + .34(\frac{33}{34}, 0, \frac{1}{34})$ and $.66(0, 1, 0) + .34(0, 1, 0)$. The second is between $.66(0, 0, 1) + .34(\frac{33}{34}, 0, \frac{1}{34})$ and $.66(0, 0, 1) + .34(0, 1, 0)$. If independence axiom holds, then either the first option should be chosen in both cases or the second option should be chosen in both.

Perhaps the simplest departure from the expected utility model that allows for the preferences as displayed in Allais' Paradox is to relax the independence axiom as follows.

Definition 5.9 (Betweenness)**.** A continuous rational preference relation $\succeq$ satisfies *betweenness* if for all $p, q \in \mathcal{L}$ and for all $\alpha \in (0, 1)$:

1. If $p \succ q$, then $\alpha p \oplus (1 - \alpha)q \succ q$.

2. If $p \sim q$, then $\alpha p \oplus (1 - \alpha)q \sim q$.

[16] In order to interpret such statements, we would need a theory of salient outcomes (zero prize) and similarity (why is .33 more similar to .34 than 2400 to 2500). Rubinstein (1988) and more recent approach by Bordalo, Gennaioli and Shleifer under the title 'Salience' (e.g. Bordalo et al. (2012)) develop these ideas.

You should observe that this gives rise to linear (and non-crossing) indifference curves in the probability simplex. Dekel (1986) gave this axiomatization and an associated representation.

Exercise 5.10. Draw an indifference map in the two-dimensional probability simplex that is compatible with the modal answers in Allais' Paradox and also with preferences satisfying Betweenness.

5.2.4 Appendix: Ambiguity

The following thought experiment is known as Ellsberg's paradox and it questions the existence of well-defined subjective probabilities. A decision maker chooses between Urn 1 and Urn 2. Urn 1 contains 50 red and 50 blue balls. Urn 2 contains x red balls and $100 - x$ blue balls (with $0 \leq x \leq 100$). A ball is picked from the urn at random and the decision maker gets EUR 10 if the ball is red. In experiments, decision makers often express a strict preference for Urn 1.

After the ball is drawn, an independent observer notes the color and put the ball back in the urn without telling its color to the decision maker. In the second task, the decision maker chooses again the urn and wins EUR 10 if a blue ball is drawn. Most decision makers still express a strict preference for Urn 1.

Do you see the paradox here? You should choose the urn with the larger likelihood of winning. If a red ball is more likely from Urn 1 than from Urn 2, then a blue ball should be more likely from Urn 2 (since all balls are either red or blue).

Ellsberg and many economists after him attribute the observed behavior to decision makers' preference for well-defined odds over more *ambiguous* probability assessments. The resulting theory takes one of two possible approaches: i) abandon subjective probabilities and evaluate uncertain prospects using Choquet's capacities that are supermodular probabilities (where for some events A, B such that $A \cap B = \emptyset$, one can have $p(A \cup B) > p(A) + p(B)$). ii) Abandon the requirement of a single subjective probability assessment and allow for a set of priors for the decision maker when making the choice. The notes by Strzalecki on decision theory on the syllabus develop this theory in Chapter 10.

5.3 Monetary Payoffs and Risk

By far the most important application of choice under risk is to the case where the consequences are monetary payoffs. This covers large parts of financial economics, consumer behavior with savings and many other important applications. Here, it will be convenient to take $X \subset \mathbb{R}$, and the idea is that $x \in X$ represents a final wealth and the decision maker has preferences over distributions of final wealth as given by the expected utility formula. We will write F for the distribution and we allow for both discrete, continuous and mixed distributions.

In general, we write the expected utility as the Stieltjes integral:

$$U(F) = \int u(x) dF(x).$$

If the distribution is discrete with probability mass function p on discrete values $\{x_k\}$, this can be written as the familiar summation:

$$U(F) = \sum_k p(x_k) u(x_k).$$

If the probability distribution is continuous with density f, we write:

$$U(F) = \int u(x) f(x) dx.$$

5.4 Risk Attitudes

Our first substantive application of the expected utility theorem is towards understanding the decision makers' attitudes towards risk. We start by finding a measure in terms of certain final wealth that is considered equally good as a risky lottery.

Definition 5.11. The *certainty equivalent* $c(F, u)$ of a lottery with distribution F for a decision maker with (Bernoulli) utility function u is defined by

$$u(c(F, u)) = \int u(x) dF(x) \tag{4}$$

We can discuss attitudes towards risk by comparing the certainty equivalents of a fixed lottery under different utility functions.

Definition 5.12. A decision maker with a utility function u is said to be *risk averse* if for all F,

$$c(F, u) \leq \int x dF(x). \tag{5}$$

Jensen's inequality implies the following result on the representation of risk averse preferences:

Proposition 5.13. A utility function u is risk averse if and only if it is concave.

Risk loving attitudes are defined with the opposite inequalities.A risk neutral decision maker is one that cares only about the expected value of the lottery. Such an agent cannot be strictly risk averse or strictly risk loving and therefore her Bernoulli utility function must be linear.

A natural question to ask is how risk aversion should be quantified. If the Bernoulli utility function is twice differentiable, then you can relate risk-aversion to negative second derivatives. Given that Bernoulli utility functions are unique only up to affine transformations, simply looking at the numerical value of the second derivative is not a good idea. If we adjust this measure by the slope of the function, we get the most used measure for risk aversion.

Definition 5.14. The *Arrow-Pratt measure of absolute risk aversion*, $r_A(x, u)$ of Bernoulli utility function u at wealth level x is given by:

$$r_A(x, u) = -\frac{u''(x)}{u'(x)}. \tag{6}$$

The following theorem shows that $r_A(x, u)$ is a good measure of risk aversion.

Proposition 5.15. The following are equivalent: i) $r_A(x, u_2) \geq r_A(x, u_1)$ for all x.

ii) There is a concave function $\psi(\cdot)$ such that $u_2(x) = \psi(u_1(x))$.

iii) $c(F, u_2) \leq c(F, u_1)$ for all $F(x)$.

Proof. We prove that i) and ii) are equivalent under the assumption that u_1 and u_2 are both twice differentiable (and therefore continuous). Then there exists a twice differentiable and increasing $\psi : \mathbb{R} \to \mathbb{R}$ such that $u_2 = \psi(u_1)$.

Differentiating by chain rule:

$$u_2'(x) = \psi'(u_1(x))u_1'(x)),$$

and differentiating again gives:

$$u_2''(x) = \psi'(u_1(x))u_1''(x) + \psi''(u_1(x))(u_1'(x))^2.$$

Dividing both sides by $u_2'(x) = \psi'(u_1(x))u_1'(x)$ gives:

$$r_A(x, u_2) = r_A(x, u_1) - \frac{\psi''(u_1(x))}{\psi'(u_1(x))}u_1'(x).$$

Thus $r_A(x, u_2) \geq r_A(x, u_1)$ for all x if and only if $\psi''(u_1) \leq 0$ for all u_1 in the domain of ψ (i.e. the range of u_1).

To see the equivalence of ii) and iii) recall that:

$$\begin{aligned} c(F, u_2) &= u_2^{-1}(\mathbb{E}_F(u_2)) \\ &= u_1^{-1}(\psi^{-1}(\mathbb{E}_F(\psi(u_1)))). \end{aligned}$$

By Jensen's inequality, ψ is concave if and only if $\mathbb{E}_F(\psi(u_1)) \leq \psi(\mathbb{E}_F(u_1))$ for all F. Thus ψ is concave if and only if:

$$\begin{aligned} c(F, u_2) &= u_1^{-1}(\psi^{-1}(\mathbb{E}_F(\psi(u_1)))) \\ &\leq u_1^{-1}(\psi^{-1}(\psi(\mathbb{E}_F(u_1)))) \\ &= u_1^{-1}(\mathbb{E}_F(u_1)) = c(F, u_1). \end{aligned}$$

□

A related concept is the measure of relative risk aversion.

Definition 5.16. The *Arrow-Pratt measure of relative risk aversion* of Bernoulli utility function u at wealth level x is given by:

$$r_R(x,u) = -\frac{xu''(x)}{u'(x)}.$$

$r^R(x,u)$ measures the attitudes towards gambles proportional to wealth. Weak empirical evidence suggests that people are risk averse, their absolute risk aversion decreases with wealth, and their relative risk aversion decreases or is constant with wealth.

Exercise 5.17. Define the prudence of a Bernoulli utility function u as:

$$P(x,u) = -\frac{u'''(x)}{u''(x)}.$$

In other words, $P(x,u)$ is the risk aversion of $-u'(x)$. What are the conditions for decreasing absolute risk aversion and decreasing relative risk aversion when expressed with $P(x,u)$ and $r_A(x,u)$?

5.4.1 Special Types of Bernoulli Utility Functions

Constant absolute risk aversion (CARA):

$$u(x) = a - be^{-\gamma x}.$$

Then $r_A(x,u) = \gamma$ for all x

Constant relative risk aversion (CRRA):

$$u(x) = a + b\frac{x^{1-\rho}}{1-\rho} \text{ for } \rho \neq 1.$$

Then $r_R(s,u) = \rho$ for all x.

Exercise 5.18. Show that in the limit as $\rho \to 1$, we get $u(x) \to \ln x$. (How do you have to specify a, b in this case?)

Example 5.19. A strictly risk averse decision maker must allocate an amount y between two identical statistically independent investment opportunities. Denote the random (gross) return on opportunity i by $(1+r_i)$ and let $f(r_i)$ denote the density function of the return. Let α be the fraction of wealth invested in opportunity 1. The final wealth of the investor is then:

$$\tilde{w}(\alpha) = (1+r_1)\alpha y + (1+r_2)(1-\alpha)y.$$

The expected utility of the investor that follows strategy α is then (by independence):

$$v(\alpha) = \int\int u((1+r_1)\alpha y + (1+r_2)(1-\alpha)y) f(r_1) f(r_2) dr_1 dr_2.$$

Notice that

$$v''(\alpha) = \int\int (r_1-r_2)^2 y^2 u''((1+r_1)\alpha y + (1+r_2)(1-\alpha)y) f(r_1) f(r_2) dr_1 dr_2 < 0$$

since the decision maker is strictly risk averse. Hence first order conditions are also sufficient for maximum.

The optimal α is found by setting

$$v'(\alpha) = \int\int (r_1-r_2) y u'((1+r_1)\alpha y + (1+r_2)(1-\alpha)y) f(r_1) f(r_2) dr_1 dr_2 = 0.$$

But then we must have:

$$\int\int r_1 u'((1+r_1)\alpha y + (1+r_2)(1-\alpha)y) f(r_1) f(r_2) dr_1 dr_2 = \int\int r_2 u'((1+r_1)\alpha y + (1+r_2)(1-\alpha)y) f(r_1) f(r_2) dr_1 dr_2.$$

Since $f(r_1) = f(r_2)$, the two sides are equal if $\alpha = \frac{1}{2}$. By strict concavity, this must be the only solution.

Example 5.20. A risk averse decision maker currently at wealth level $290,000 turns down a bet that gives winnings $11 with probability .5 and loses $10 with probability .5 at levels of wealth, $w \leq \$300,000$. How large would the winnings have to be in a bet that loses $200 with probability .5 for the decision maker to accept the bet.

Since we are not given the exact form of the Bernoulli utility function, we can get at best a lower bound for the amount.

If a decision maker behaves according to expected utility theory and is risk averse, his Bernoulli utility function, u, satisfies

$$u(w+11) - u(w) \; < \; u(w) - u(w-10).$$

and thus the marginal utility function $mu\,(w)$ satisfies

$$\begin{aligned} mu(w+11) \;&\leq\; \frac{u(w+11)-u(w)}{11} \\ &<\; \frac{10}{11}\frac{u\,(w)-u\,(w-10)}{10} \leq \frac{10}{11} mu\,(w-10)\,. \end{aligned}$$

Thus, marginal utility falls at a faster rate than that of a geometrical sequence. It is easily calculated that in order to risk $200 with probability .5, the winnings must be at least $\$12,210,000$. To get this figure, you should evaluate $u\,(300,000)$ using the inequalities above and then you should get an upper bound for $mu\,(300,000)$ and extrapolate under the assumption of linear utility for $w \geq 300,000$.

Exercise 5.21. Show that CARA expected utility functions are bounded from above and unbounded from below. Estimate the coefficient of risk aversion for a decision maker with a CARA utility that refuses $p = (.5, .5)$ on $X = \{11, -10\}$. Suppose that her initial wealth is 500. Find the smallest loss y such that a 50-50 bet on winning x or losing y is rejected for all x.

5.5 Comparing Risks

Up to now, we have been discussing properties of Bernoulli utility functions that allow us to identify some decision makers as more risk-averse than others. Now we change the perspective. We want to classify risks in a way that is valid for a wide class of decision makers.

We start with a very strong notion for dominance. It applies to all monotone increasing Bernoulli utility functions. The second class evaluates risks in a way that is valid for all risk-averse decision makers. We assume that $x \in [\underline{x}, \overline{x}]$ and that u is differentiable throughout.

5.5.1 First order stochastic dominance

For this notion, the utility functions of interest are all increasing functions, i.e. we consider

$$\Omega = \{u(x) : u'(x) \geq 0\}.$$

Definition 5.22. A distribution $F_1(x)$ *first-order stochastically dominates (FOSD)* distribution $F_2(x)$ if all decision makers with an increasing Bernoulli utility function prefer the lottery with distribution $F_1(x)$ to $F_2(x)$.

Proposition 5.23. $F_1(x)$ first-order stochastically dominates $F_2(x)$ if and only if

$$F_1(x) \leq F_2(x) \text{ for all } x \in [\underline{x}, \overline{x}].$$

Proof.

$$\int_{\underline{x}}^{\overline{x}} u(x)\, dF_1(x) - \int_{\underline{x}}^{\overline{x}} u(x)\, dF_2(x)$$

$$= \Big|_{\underline{x}}^{\overline{x}} u(x)(F_1(x) - F_2(x)) - \int_{\underline{x}}^{\overline{x}} u'(x)(F_1(x) - F_2(x))dx.$$

by integration by parts. The first term is 0 and the second is positive for all u such that $u'(x) \geq 0$ if and only if

$$F_1(x) \leq F_2(x) \text{ for all } x \in [\underline{x}, \overline{x}].$$

□

You should note that if two lotteries are ranked in the first-order stochastic dominance order, then the dominant one is better for all decision makers with increasing preferences in prizes. This includes not only risk-loving preferences but also many classes of non-expected utility preferences. For example cumulative prospect theory preferences discussed in the next section fall into this class.

A final recent observation is the following. Suppose that a decision maker chooses between two lotteries (i.e. random variables) X and Y in a situation where she is exposed to large independent background noise Z so that her final wealth is $X + Z$ or $Y + Z$ depending on her choice. Then Mu et al. (2023) show that if $\mathbb{E}X > \mathbb{E}Y$, then $X+Z$ first-order stochastically dominates $Y + Z$. I.e. all decision makers with monotone payoffs should choose X over Y. You should check the paper for the exact conditions of what 'large background risk' means in the paper.

5.5.2 Second order stochastic dominance

For this notion, the relevant class of Bernoulli utility functions is given by $\Omega' = \{u(x) : u''(x) \leq 0\}$, i.e. the comparisons should hold for all risk-averse decision makers. Note that we are not requiring u to be increasing here.

Definition 5.24. A distribution $dF_1(x)$ *second-order stochastically dominates (SOSD)* $dF_2(x)$ if

$$\int u(x)\, dF_1(x) \geq \int u(x)\, dF_2(x) \text{ for all } u(x) \in \Omega'. \tag{7}$$

We have the following characterization for SOSD:

Proposition 5.25. $dF_1(x)$ second-order stochastically dominates $dF_2(x)$ if and only if

$$\int_{\underline{x}}^{x} F_1(x)\, dx \leq \int_{\underline{x}}^{x} F_2(x)\, dx \text{ for all } x \in [\underline{x}, \overline{x}],$$

and

$$\int_{\underline{x}}^{\overline{x}} F_1(x)\,dx \leq \int_{\underline{x}}^{\overline{x}} F_2(x)\,dx$$

.

Proof. Observe first that if F_1 SOSD F_2, we must have $\mathbb{E}_{F_1}x = \mathbb{E}_{F_2}x$ since linear increasing and decreasing functions are in Ω' if F_1 is to dominate F_2.

Integrate both sides of 7 by parts twice, and cancel equal terms on both sides to get:

$$-\Big|_{\underline{x}}^{\overline{x}} u'(x)\int_{\underline{x}}^{x} F_1(s)ds + \int_{\underline{x}}^{\overline{x}} u''(x)\int_{\underline{x}}^{x} F_1(s)dsdx$$

$$\geq -\Big|_{\underline{x}}^{\overline{x}} u'(x)\int_{\underline{x}}^{x} F_2(s)ds + \int_{\underline{x}}^{\overline{x}} u''(x)\int_{\underline{x}}^{x} F_2(s)dsdx.$$

The first terms on the two sides of the inequality are equal since $\mathbb{E}_{F_1}x = \mathbb{E}_{F_2}x$. Since $u'' \leq 0$, the lhs is larger than the rhs for all $u \in \Omega'$ if and only if

$$\int_{\underline{x}}^{x} F_1(x)\,dx \leq \int_{\underline{x}}^{x} F_2(x)\,dx \text{ for all } x \in [\underline{x}, \overline{x}].$$

□

Rothschild and Stiglitz (1970) coined the term *mean-preserving spread* to denote a distribution that is obtained from another one by splitting probability mass at any point to separate points in such a way that the mean of the random variable remains fixed.

Another characterization of SOSD can be given in terms of the random variables, i.e. the lotteries themselves. A lottery $\tilde{x}$ second-order stochastically dominates $\tilde{y}$ its distribution F_x second order stochastically dominates F_y.

Theorem 5.26. Random variable $\tilde{x}$ second order stochastically dominates random variable $\tilde{y}$ if and only if we can write

$$\tilde{y} = \tilde{x} + \tilde{z},$$

where $E[\tilde{z}\,|\tilde{x}] = 0$.

We say that in this case, $\tilde{y}$ is $\tilde{x}$ plus 'noise'.

5.5.3 Illustrations

Example 5.27. Consider a maximization problem where the utility depends on the realization of a random variable as well as the chosen value of a control variable. Let $u(a, x)$ be the utility function, where a is the control variable and x is the uncertain variable. In many applications, e.g. when a is the level of investment in capacity, and $u(a, x)$ is the level of output at capacity a and realization x, the following assumptions are natural to make: $u_a(a, x) > 0$ and $u_{aa}(a, x) < 0$.[17] Let a denote the control variable, and let $F(x; r)$ be the distribution of the random variable with risk parameter r. Assume that higher values of r represent risks that are second-order stochastically dominated by risks with lower r. The problem is then to

$$\max_a \int_0^1 u(a, x)\, dF(x; r) = \max_a \int_0^1 u(a, x)\, f(x; r)\, dx,$$

where $f(x; r)$ is the density function of the random variable. The first derivative of the expected utility from choice a is given by

$$v'(a) = \int_0^1 u_a(a, x)\, f(x; r)\, dx.$$

Observe that

$$v''(a) < 0 \text{ since } u_{aa}(a; x) < 0.$$

If we are interested in the comparative statics of the control variable in the risk parameter, we need to apply the results that we had for second

[17] I am using subscripts on u to denote partial derivatives.

order stochastic dominance above. Define $a(r)$ to be the optimal choice for risk level r. Then

$$v'(a(r)) = \int_0^1 u_a(a,x)\, f(x;r)\, dx = 0.$$

We know that if $u_a(a,x)$ is concave in x, then

$$\int_0^1 u_a(a(r),x)\, f(x;r')\, dx \leq 0 \text{ for } r' \geq r.$$

This simply reflects the fact that the expected utility for concave functions is lower from risks that are second order stochastically dominated (i.e. the proposition of the previous section). Since $v(a)$ is concave, this implies that $a(r') < a(r)$. Hence the optimal action is decreasing in r. The opposite conclusion follows if $u_a(a,x)$ is convex in x. To summarize, we have shown the following proposition.

Proposition 5.28. Let the utility from action a and outcome x is given by a function $u(a,x)$ and the distribution of the risk is given by $F(x;r)$, where r is a parameter of increasing risk in the sense of second order stochastic dominance. The optimal choice $a(r)$ as a function of the risk is
i) decreasing if $u_{axx}(a,x) \leq 0$,
ii) increasing if $u_{axx}(a,x) \geq 0$.

5.6 Applications of the Expected Utility Theorem

5.6.1 Standard Portfolio Choice

Consider a risk averse decision maker with initial wealth w_0. Her decision problem is to choose how much to invest in safe versus risky assets? We do not allow short sales so since the decision maker does not hold assets initially, she can only buy them. Let $(1+r)$ denote the riskless gross return on a safe asset, and let $(1+\widetilde{x})$ denote the random return on the risky investment. Denote the amount of risky investment by $0 \leq \alpha \leq w_0$, and the safe investment by $(w_0 - \alpha)$.

The final wealth of the decision maker is given by:

$$(w_0-\alpha)(1+r)+\alpha(1+\widetilde{x})=w_0(1+r)+\alpha(\widetilde{x}-r).$$

Assume that the decision maker has a strictly concave, strictly increasing twice differentiable Bernoulli utility function $u(w)$. Her expected utility from a risky investment α is given by:

$$v(\alpha)=Eu(w_0(1+r)+\alpha(\widetilde{x}-r)).$$

The value of her investment $v(\alpha)$ is a strictly concave function of α if $\Pr(\widetilde{x}=r)<1$ since

$$v''(\alpha)=E\left((\widetilde{x}-r)^2 u''(w_0(1+r)+\alpha(\widetilde{x}-r))\right).$$

As a result, the first-order condition for optimal α is also a sufficient condition for maximum. The first order condition for interior solutions (i.e. for solutions where $0<\alpha<w_0$) is given by:

$$v'(\alpha)=E(\widetilde{x}-r)u'(w_0(1+r)+\alpha(\widetilde{x}-r))=0.$$

For $\alpha=0$, it must be that :

$$v'(0)=E(\widetilde{x}-r)u'(w_0(1+r))\leq 0.$$

Since $u'(w_0(1+r))$ is independent of $\widetilde{x}$, the above condition is equivalent to

$$u'(w_0(1+r))E(\widetilde{x}-r)\leq 0.$$

Hence a necessary condition for no risky investments is that the expected value of the investment be no larger than the safe return. This is also a sufficient condition (exercise: why?). We conclude that all decision makers, risk averse or not, invest some positive amount in risky assets if their expected return is larger than the safe rate. Another way of phrasing this observation is that all expected utility decision makers are approximately risk-neutral for small bets.

Comparing Risk Attitudes in Portfolio Choice Consider next two risk averse decision makers with utility functions $u_1$and u_2. Suppose that u_1 is more risk averse than u_2. Then $u_1(x) = \phi(u_2(x))$ for some concave function ϕ.

We want to see how the optimal portfolio choices of u_1 and u_2 can be compared. Since the utility functions are assumed to be concave, the first-order conditions are also sufficient for optimality.

Denote the optimal risky investments for the two utility functions by α_1 and α_2 respectively. From the first order condition for u_2, we have:

$$v_2'(\alpha_2) = E(\widetilde{x} - r)\, u_2'(w_0(1+r) + \alpha_2(\widetilde{x} - r)) = 0. \tag{8}$$

To see how the optimal risky investment of u_1 relates to α_2, we evaluate the derivative of $v_1(\cdot)$ at $\alpha = \alpha_2$.

$$v_1'(\alpha_2) = \frac{d}{d\alpha} E\phi(u_2(w_0(1+r) + \alpha(\widetilde{x} - r)))\,|_{\alpha=\alpha_2}$$
$$= E(\widetilde{x} - r)\, \phi'(u_2(w_0(1+r) + \alpha_2(\widetilde{x} - r)))\, u_2'(w_0(1+r) + \alpha_2(\widetilde{x} - r)).$$

Since $\phi'' \leq 0$, we know that

$$(\widetilde{x} - r)\, \phi'(u_2(w_0(1+r) + \alpha_2(\widetilde{x} - r))) \leq (\widetilde{x} - r)\, \phi'(u_2(w_0(1+r)))$$

for all $\widetilde{x}$.

To see this note that for $\widetilde{x} < r$,

$$\phi'(u_2(w_0(1+r) + \alpha_2(\widetilde{x} - r))) \geq \phi'(u_2(w_0(1+r))),$$

by the concavity of ϕ, and similarly for $\widetilde{x} > r$,

$$\phi'(u_2(w_0(1+r) + \alpha_2(\widetilde{x} - r))) \leq \phi'(u_2(w_0(1+r))),$$

and hence the claim follows. But then we know that

$$v_1'(\alpha_2) \leq E(\widetilde{x} - r)\, \phi'(u_2(w_0(1+r)))\, u_2'(w_0(1+r) + \alpha_2(\widetilde{x} - r))$$
$$= \phi'(u_2(w_0(1+r)))\, E(\widetilde{x} - r)\, u_2'(w_0(1+r) + \alpha_2(\widetilde{x} - r)) = 0,$$

where the last equality follows from 8. Thus by the concavity of $v_1(\alpha)$, we know that $\alpha_1 \leq \alpha_2$.

Proposition 5.29. If u_1 is more risk averse than u_2, then $\alpha_1 \leq \alpha_2$ in the standard portfolio problem.

This proposition also yields an immediate corollary for risky investment as a function of initial wealth.

Corollary 5.30. If u satisfies decreasing absolute risk aversion, then $\alpha(w_0) \leq \alpha(w_0')$ whenever $w_0 < w_0'$.

Proof. Take $u_2(z) = u(z)$ and $u_1(z) = u(z-k)$ and apply the previous theorem. □

5.6.2 Consumption and Savings

We start with the simplest deterministic two-period model, and derive conclusions for optimal savings and consumption. We assume additively separable utility function over the two periods. In other words, the consumer has a separate Bernoulli utility function for consumption in each period $t = 0, 1$.

The consumer receives wealth w_0 and w_1 respectively in the two periods. She can borrow and lend as she wishes at the risk free rate r. If we let s denote the savings by the consumer, then her optimization problem can be written as

$$\max_s \; u_0(w_0 - s) + u_1(w_1 + s(1+r)).$$

Observe that we can allow for negative saving (i.e. borrowing) in this model, but we require that consumption be positive in both periods (i.e. $s \leq w_0$). Assume throughout that $u_t(\cdot)$ are strictly concave and twice continuously differentiable for $t = 0, 1$.

Hence if we let

$$v(s) = u_0(w_0 - s) + u_1(w_1 + s(1+r)),$$

we see immediately that $v''(s) < 0$. This allows us again to locate optimal savings levels from the first order conditions.

The optimal level of savings s^* is characterized by

$$v'(s^*) = -u_0'(w_0 - s^*) + (1+r)\, u_1'(w_1 + s^*(1+r)) = 0.$$

If $u_0 = u_1 = u$ and $r = 0$, we see the most clearly how savings are used to smooth consumption across periods. From

$$u'(w_0 - s^*) = u'(w_1 + s^*),$$

we conclude by the strict concavity of u that

$$w_0 - s^* = w_1 + s^*.$$

Hence the consumption levels in the two periods are identical.The other main motive of saving is to increase wealth. This effect can obviously only be seen when $r > 0$. Again in the case where $u_0 = u_1 = u$, we get:

$$u'(w_0 - s^*) = (1+r)\, u'(w_1 + s^*(1+r)).$$

By the concavity of u, we see that consumption in the second period is larger (since the marginal utility is lower) than in the first period. Hence the consumer is willing to sacrifice some of the consumption smoothing for increases in wealth.

Finally, we can totally differentiate the first-order condition with respect to s and w_i to get

$$\frac{ds^*}{dw_0} = \frac{u_0''(w_0 - s^*)}{\left[u_0''(w_0 - s^*) + (1+r)^2 u_1''(w_1 + s^*(1+r))\right]} > 0,$$

$$\frac{ds^*}{dw_1} = \frac{-u_1''(w_1 + s^*(1+r))(1+r)}{\left[u_0''(w_0 - s^*) + (1+r)^2 u_1''(w_1 + s^*(1+r))\right]} < 0.$$

Hence an increase in the first period income increases savings, and an increase in the second period income decreases savings. With these preliminaries in place, we can start the analysis of the optimal savings problem in a world of uncertainty.

The first question that we ask is whether the optimal savings are larger in a model where the second period income is random than in the deterministic model.

Definition 5.31. A utility function is *prudent* if adding an uninsurable zero mean risk to the second period income increases the savings.

To characterize prudent utility functions, let $\widetilde{w}_1 = w_1 + \widetilde{x}$, where $\widetilde{x}$ is assumed to be uninsurable and $E\widetilde{x} = 0$. Denote the new expected utility from savings s by:

$$V(s) = u_0(w_0 - s) + Eu_1(w_1 + s(1+r) + \widetilde{x}).$$

We note that as before, $V(s)$ inherits the curvature of the u_i functions. We analyze comparative static questions by evaluating the derivative of $V(s)$ at point s^* such that $v'(s^*) = 0$, i.e. at the optimal savings level of the deterministic model.

Observe that $V'(s^*) \geq 0$ if

$$Eu_1'(w_1 + s^*(1+r) + \widetilde{x}) \geq u_1'(w_1 + s^*(1+r)). \tag{9}$$

Notice that on the left hand side of the inequality, we have the expected utility from a random variable. On the right hand side, we have the utility from the expected value of the random variable.

This is exactly the definition of a risk loving utility function since w_1 and $\widetilde{x}$ are arbitrary. As risk loving functions are convex, we deduce that 9 holds for all w_1 and $\widetilde{x}$ if and only if u_1' is convex. Hence we have proved the following proposition.

Proposition 5.32. A utility function is prudent if and only if u_1' is convex.

From this point on, we could develop a theory for comparing prudence of different individuals or the prudence of a given individual at various wealth levels. Much of this theory has been done by Miles Kimball, and the central concept for the analysis is the coefficient of absolute prudence:

$$P(w) = \frac{-u'''(w)}{u''(w)}.$$

We conclude this section on precautionary savings by recalling the derivation for decreasing absolute risk aversion from an exercise in the previous section:

$$\frac{d}{dw} r^A(w) = r^A(w) \left[r^A(w) - P(w) \right].$$

Hence there are two arguments for believing in the prevalence of prudent utility functions. First of all, there is direct econometric evidence on the savings behavior of individuals with various degrees of uninsurable risk positions. Second, there is overwhelming empirical support for decreasing absolute risk aversion. As the formula above indicates, DARA is only possible for prudent utility functions.

5.7 Appendix: Blackwell's Theorem

Consider the single-agent decision problem, where the agent chooses the optimal action $a \in A$ to maximize her expected utility. The utility depends on the chosen action and on the state of nature $\theta \in \Theta = \{\theta_1, ..., \theta_K\}$ so that the optimal action under full information depends on the state. Assume also that the (state dependent) utility function $u(a, \theta)$ is continuous in a for all θ and A is compact. Let p_k denote the prior probability of the event $\{\theta = \theta_k\}$, and $p = (p_1, ..., p_K)$. We assume w.l.o.g. that $p_k > 0$ for all k. The decision maker's problem is then to

$$\max_a \sum_{k=1}^{N} p_k u(a, \theta_k).$$

Let $a(p)$ denote the maximizer in the above program, and let

$$V(p) = \sum_{k=1}^{N} p_k u(a(p), \theta_k)$$

be the value function of the program. As we argued in Section 3.4, $V(p)$ is convex in p.

If there are two states, then $p \in [0,1]$. Therefore, SOSD (with signs changed since we are dealing now with convex functions) gives a characterization for distributions on probabilities that are good for the decision maker. A statistical experiment induces a distribution F on probability distributions that must satisfy the Bayes' plausibility constraint: $\mathbb{E}_F(p') = p$, where p' is a posterior induced by the experiment. We can ask when experiment with distribution F is preferred to an experiment G for all decision makers (i.e. for all utility functions) and for all priors.

The best the decision maker could hope for is to know the true state resulting in $p' \in \{0,1\}$. Clearly the perfectly revealing statistical experiment is the best for the decision maker (since then she can take optimal actions state by state). Loosely speaking, an experiment is a random variable whose outcome is correlated with the true state of the world. After seeing the outcome in the experiment, the decision maker updates her beliefs on the state and then chooses the optimal action. The above reasoning tells us that one experiment is better than a second experiment if the posterior belief resulting from the first is second-order stochastically dominated by the belief resulting from the second. Or to put it slightly differently, if the posterior from the first is a mean preserving spread of the posterior from the second.

In future courses on Economics of Information, we will see Blackwell's theorem in action in more general settings.

6 Advanced Topics

6.1 Probability Weighting

There is a lot of experimental evidence that decision makers do not take probabilities at face values but interpret them subjectively. In particular, there is evidence that decision makers overweight small probabilities and underweight high probabilities. How could one model this?

A naive approach to this would be to just distort the probabilities in the Expected Utility formula by some function $\psi : [0,1] \to [0,1]$ so that we would have:

$$\mathbb{E}U = \sum_k \psi(p(x_k))u(x_k).$$

The problem with this formulation is that the distorted $\psi(p(x_k))$ do not necessarily sum up to 1. This leads to undesirable properties for the distorted formula. For example, such preferences do not satisfy monotonicity in first-order stochastic dominance order on risks (even for increasing u and ψ).

To avoid such problems, a different approach is needed. Even though the following treatment is not the original one given by Kahnemann and Tversky (1979) or Quiggin (1982), it is perhaps the easiest to digest. Recall the alternative formulation of the expected utility formula in terms of the quantiles:

$$\mathbb{E}_F U = \int_0^1 u(F^{-1}(s)ds = \int_0^1 u(f_X(s))ds,$$

where $X \sim F(\cdot)$, and $f_X(s) = F^{-1}(s)$. This formulation gives equal weight to all quantiles of the random variable X. We can distort the probabilities by shifting to another probability measure $\pi(s)$ on $[0,1]$ to get the rank-dependent utility formula with distortion function π:

$$\int_0^1 u(f_X(s))d\pi(s).$$

A typical finding in experimental studies trying to estimate the probability distortion function is that it is concave for small s and convex for large s. This corresponds to the exaggerating the importance of events with a small probability and downplaying the value from probable events.[18] If you take the distortion to be the point mass $\delta_{0.05} = \pi$, then you get the value at risk (VaR) evaluation that is used in many financial regulatory settings. Convex probability weighting functions help explain the choices leading to Allais' Paradox.

There are a few axiomatizations of preferences leading to the rank-dependent utility formula. Perhaps the most used one is by Yaari (1987) that results in a linear u combined with the probability weighting. The key axiom in this approach is an independence axiom, but now stated in terms of the quantile functions. I should note that Yaari's formulation has been recently used very successfully in economic applications to insurance markets and to screening. We saw in the example on standard portfolio choice that decision makers with (differentiable) expected utility preferences are approximately risk-neutral for small gambles. Yaari's preferences allow for first-order risk-aversion (for EU, risk-aversion is only reflected in the second-order term of the Taylor approximation of u). This can be viewed as a positive feature of Yaari's model.

On the negative side, the lack of local risk-neutrality implies that if $\pi(s) \neq s$, then one can always find small negative expected value gambles that the decision makers would accept making them susceptible for money-pump arguments. See Ebert and Strack (2015).

6.2 Choosing Menus

Kreps (1979) proposed a model where the preferences of a decision maker are on menus of lotteries (i.e. sets of lotteries) rather than just on individ-

[18]In the applied literature on risky choice, a recurring empirical observation is the longshot-favorite bias in betting. On the race-track, favorites have a higher actuarial rate of return than longshots.

ual lotteries. His motivation was to analyze issues such as preference for flexibility. If some future circumstances are unknown at an initial stage, it may be a good idea to have many possible choices at a future stage when uncertainty has been resolved.

Kreps modeled the situation as follows. There is a set of possible states S in period $t = 1$ and at $t = 0$, you have a preference for choice menus of alternatives for period $t = 1$. A simple example would be that in $t = 0$, you need to pack either rubber boots or sandals. The state would be the weather in $t = 1$: either rain or shine. If you do not know $s \in S$, you might be indifferent between boots and sandals, but have a preference for the menu containing both boots and sandals.

Let $A_1, B_1, C_1, ... \subset X_1$ denote the menus for period $t = 1$ containing objects $x_1 \in X_1$. Let $p(s)$ be the probability assignment for each $s \in S$. The $t = 0$ preferences $\succeq_0$ have an *option value representation* u_0 if there is a state-dependent utility function $u_1 : X_1 \times S \to \mathbb{R}$ such that for all $A_1 \subset X_1$:

$$u_0(A_1) = \sum_{s \in S} p(s) \max_{x_1 \in A_1} u_1(x_1, s).$$

Kreps showed that the preference $\succeq_0$ has an option value representation u_0 if and only if $\succeq_0$ satisfies the following two axioms.

Axiom 6.1 (Preference for flexibility)**.**
If $B_1 \subset A_1$, then $A_1 \succeq_0 B_1$.

Axiom 6.2 (Modularity)**.**
If $A_1 \sim_0 A_1 \cup B_1$ then for all C_1, we have $A_1 \cup C_1 \sim_0 A_1 \cup B_1 \cup C_1$.

If one wants to extend Kreps' model to include lotteries in the choice set, then one must impose the independence axiom and a continuity axiom to get the equivalent representation. While this is a nice theorem, it is not really that great as a representation since it only guarantees the existence of a state-dependent utility and utilities of this form are not easy to identify.

Gul and Pesendorfer (2001) took the opposite view on added options. Rather than flexibility, they wanted to model temptation and the related desire to commitment. Even if you end up choosing a healthy option on a menu in a restaurant, you may suffer disutility if you have to refuse tempting less healthy items. Because of this consideration, a smaller menu might sometimes be preferable to a larger one. Their construction is directly in terms of menus of lotteries so a continuity axiom on the preferences and the independence axiom are assumed throughout. The substantive axiom in Gul and Pesendorfer (2001) that yields a very nice representation is the following.

Axiom 6.3 (Set Betweenness)**.**
$A_1 \succeq_0 B_1$ implies $A_1 \succeq_0 A_1 \cup B_1 \succeq_0 B_1$.

Theorem 6.4. The rational preference $\succeq_0$ on menus satisfies Independence, Continuity and Set Betweenness if and only if there is a representation u, v:

$$u_0(A_1) = \max_{x_1 \in A_1} u(x_1) + v(x_1) - \max_{y_1 \in A_1} v(y_1).$$

The representation here is both very tight (just two Bernoulli functions) and yields a nice representation. $u(x_1)$ is the final utility resulting from the choice x_1 while $\max_{y_1 \in A_1} v(y_1) - v(x_1)$ denotes the loss in utility from having to refuse the most tempting option y_1.

Gul and Pesendorfer (2004) extend this model to a dynamic decision making framework that allows them to discuss many economic problem including the savings problems that have motivated much of the literature on commitment and temptation.

6.3 Behavioral Models: Time-Inconsistent Preferences, Reference Dependent Preferences

6.3.1 Time-Inconsistent Preferences

In Behavioral Economics, models of rational decision making have been augmented by additional psychological considerations to expand the scope of the theories. A prime example of this is the theory of time-inconsistent decision making.

Key ideas in this literature are the following: the present has a heightened importance in the decision making process. Today versus tomorrow is a different trade-off than 30 days from now versus 31 days from now. This leads to a desire to have pleasant experiences right away and to delay unpleasant ones. A prime example of this is procrastination, i.e. the tendency to postpone boring duties. Time-inconsistency shows up in the following thought experiment. Would you do a necessary unpleasant task right away rather than let it linger for two weeks. Maybe it would be better to do it right away, but even better to do it tomorrow. If a decision maker is not aware of her future tendency to reason in the same way tomorrow as she does today, she may procrastinate with the task today in the hopes that it will be done tomorrow. But once tomorrow is here, the reasoning process will be the same again and the task will be kicked further into the future.

The most common way of modeling such preferences is to have an exponential discount factor δ applied as in the time consistent decision models, and an additional factor β applied uniformly (not exponentially) to all future periods. In this $\beta - \delta$ framework, factor β captures the *present bias* that leads to inconsistent choice.

Example 6.5. You have money for exactly 1 movie ticket and 4 days over which you might use it. The utilities from seeing the movies (on the date of seeing it) are $u_1 = 2, u_2 = 3, u_3 = 5, u_4 = 9$. At each t, you decide which $t' \geq t$ to choose for you movie. Suppose $\delta = 1$ (since the time delay is small) and $\beta = \frac{1}{2}$ so that the present bias is important. If you are unaware of your time-inconsistency, you decide in $t = 1$ to wait until $t = 4$ since $\frac{9}{2} > 2$. You do the same in $t = 2$, but in $t = 3$ you choose to go to the movie since $\frac{9}{2} < 5$. If you are aware of your inconsistency, you realize that in $t = 3$, you will not wait. Therefore in $t = 2$ you will not wait either (since $3 > \frac{5}{2}$) and therefore in $t = 1$ you go to the movie right away. We conclude that being sophisticated about your own future period will not necessarily help.

Prime examples of applications for time-inconsistent preferences are: i) Commitment to savings behavior (I would like to save for retirement, but I am afraid that if I leave money in my bank account, I will use it tomorrow for a luxury vacation. As a result, I commit to an account with no withdrawal possibilities.) ii) Choosing a restaurant without tempting options. iii) Gym membership with the idea of committing to exercising more.

In order to have a proper treatment of this topic, we should really frame the model as a game between current self and future selves. In this course, we will not deal with game-theoretic models. Note that Gul and Pesendorfer (2001) presented in the previous section gets at the same economic questions within a single decision maker framework.

6.3.2 Reference-dependent Utility

One of the key elements in Kahneman and Tversky's Prospect Theory is the idea that losses are more important than winnings to decision makers in risky situations. In other words, your current wealth is a reference point and you compare any lotteries against this reference point with different weightings for losses and gains.

It is not clear what the appropriate reference point should be in all cases, and how it should depend on the actions chosen by the decision maker. Köszegi and Rabin (2007) formalize decision making with endogenous reference points. The idea is the following: You choose amongst a set of lotteries $p, q, r, s, ...$. If p is your tentative choice, then you evaluate gains and losses from other lotteries relative to p. Your actual choice is any lottery with the property that it performs at least as well as any other lottery when compared to itself. Again this is really a game-theoretic notion of equilibrium determination for a reference point and choice.

This game-theoretic notion of reference dependent choice has a tight connection to rank-dependent utility models as explained in Masatlioglu and Raymond (2016).

6.4 Stochastic Choice

In Section 1, we discussed the possibility of having choice correspondences over choice sets $A \in X$ rather than choice functions where the decision maker chooses a subset from the available options. One way to interpret choice correspondences is that it reports for each choice set those items that are chosen with a positive probability. Stochastic choice is the part of choice theory that focuses on this interpretation.

The central object of study in stochastic choice is the *stochastic choice function* $\rho(x, A)$ that gives the probability at which x is chosen from choice set A for each $x \in A$ and all A. If A has N elements, we can view $\rho(\cdot, A)$ as a probability vector, i.e. an element in the $N-1$-dimensional simplex. These are the probabilities that an outsider analyst would assign to choices made by the decision maker. The true choice may be deterministic, but dependent on factors not observable to the analyst.

There are a number of reasons to study random choice. Individual choice behavior appears to be stochastic in experiments (see Strzalecki's Stochastic Choice Theory Section 1.3 for some references). Population level heterogeneity is reflected in fractions of population choosing different options and these fractions can be interpreted as the choice probabilities. Other reasons include learning, direction of attention in a random manner, errors in choices and deliberate randomizations. By far the most widely studied class of models is the *random utility model* where each individual decision maker is subject to a random utility shock.[19]

In this approach, utility is a random vector on a probability space $(\Omega, (F), \mathbb{P})$, i.e. we write $\tilde{U} : \Omega \to \mathbb{R}^X$, and we say that the decision maker chooses

[19] There are a number of ways to formalize the random utility model, but this is enough for our purposes here.

$x \in A$ whenever $\tilde{U}_x(\omega) > \tilde{U}_y(\omega)$ for all $y \neq x \in A$. We write the event that x is chosen in A as:

$$c(x, A) = \{\omega | \tilde{U}_x(\omega) > \tilde{U}_y(\omega) \text{ for all } y \neq x \in A\}.$$

Definition 6.6. We say that $\rho(x, A)$ has a *random utility representation* if there exists a random variable $\tilde{U} : \Omega \to \mathbb{R}^X$ such that $\rho(x, A) = \mathbb{P}(c(x, A)$ for all $A \subset X$ and all $x \in A$.

The simplest random utility model is the *additive random utility model ARU* where the decision maker maximizes $\tilde{U}(x) = v(x) + \tilde{\epsilon}(x)$. Here v is a deterministic base utility function and $\tilde{\epsilon}$ is a random perturbation. The key is that while $\tilde{\epsilon}$ is observable to the decision maker, it is not observable to the analyst. As long as X is finite, there is no loss of generality in concentrating on ARU rather than more general random utility models (i.e. the same stochastic choice functions can be rationalized in both approaches).[20]

ARU with a particular error term distribution is the work horse of discrete choice econometrics. The *logit-model* results from assuming that $\tilde{\epsilon}(x)$ are i.i.d. Type I Extreme Value random variables (also known as the Gumbel distribution). This is the most commonly used formulation in empirical IO, consumer theory etc. Probit results from Normally distributed error terms. Versions of ARU have also been used in game theory to represent smooth best responses.

It is normally assumed that $\rho(x, A) > 0$ for all $x \in A$. This is not particularly restrictive and in any case, this assumption would be impossible to falsify in a finite sample. The main axiom in stochastic choice is *regularity*.

Axiom 6.7 (Regularity)**.** The stochastic choice function $\rho(x, A)$ is regular if $A \subset B$ implies that $\rho(x, A) \geq \rho(x, B)$.

Note that this has a resemblance to Sen's α in the deterministic choice theory. Block and Marschak showed that if $\rho(x, A)$ has a random utility

[20]This is true only when the utility functions are chosen in a flexible manner. If the utility functions are restricted to a parametric class, the two approaches differ.

representation, then it is regular. The converse is true only if X has at most three elements.

Even though regularity seems a natural axiom to impose, there are reasons why it might fail. Choice overload may lead to an overall smaller probability of choosing anything. Decoy altenatives may boost the probability of a given alternative to be chosen just like in the deterministic choice part.

The logit or Luce model yields a very tractable formula for the choice probabilities:

$$\rho(x, A) = \frac{e^{v(x)}}{\sum_{y \in A} e^{v(y)}}.$$

This formulation is great for its simplicity, but leads to problems. An old observation is that by splitting an alternative into two equivalent parts, the probability of choice changes in Luce's formula (train vs bus is different from train vs red bus vs blue bus, and this is not plausible).

The remedies for this include nested logit models where the sequence of decisions made by the decision maker is formulated explicitly (so that train vs bus is chosen first and blue bus versus red bus in the second stage). Obviously such procedures are quite context-dependent.

The book by Stralecki on the syllabus is an excellent and very accessible introduction to a wide variety of stochastic choice models. Any student interested in pursuing connections between theory and empirical work is encouraged to study this topic (for IO, experimental, marketing applications etc.).