Core Microeconomic Theory

Kevin J. Mumford[1]

Econ 202N
Stanford University
Autumn 2004

[1] These notes are an edited and condensed version of the notes used by Paul Milgrom in 2003. The first version was created by Antonio Rangel in 2000. This original version was edited and expanded by Jonathan Levin in 2001 and by Paul Milgrom in 2002 and 2003. In addition, these notes draw on a set of notes created by Felix Kubler in 2002, *Microeconomic Theory* by Mas-Colell, Whinston, and Green, *A Course in Microeconomic Theory* by Kreps, and advice from Jonathan Levin and Adam Cagliarini.

Contents

Part I

Choice Theory

Chapter 1

Individual Decision Making

Individual decision-making forms the basis for nearly all of microeconomic analysis. These notes outline the "standard" economic model of *rational choice* in decision-making. In the standard view, rational choice is defined to mean the process of determining what options are available and then choosing the one that is most *preferred*.

Rational Choice

Question: What do we mean by *preferred*?

1.1 Preferences

Rational choice theory starts with the idea that individuals have preferences and choose according to those. Our first task is to formalize what that means and precisely what it implies about the pattern of decisions we should observe.

Let X be a set of possible choices. In consumer choice models, one might specify that $X \subset \mathbb{R}^n$, meaning for instance that there are n different goods (tacos, tortilla chips, salsa, etc..) and if $x \in X$, then $x = (x_1, ..., x_n)$ specifies

Choice Set

quantities of each type of good. In general, the abstractness of the choice set X allows enormous flexibility in adapting the model to various applications.

Preference Relation

Now consider an economic agent. We summarize the agent's preferences over the set X in a *preference relation* which we denote by $\succsim$. Specifically we use $\succsim$ to denote the agent's weak preference and we read $x \succsim y$ as "x is at least as good as y". From this, we can give two other definitions, the *strict preference relation* and the *indifferent preference relation*:

Definition: Strict Preference Relation $x \succ y$ iff $x \succsim y$ but not $y \succsim x$.

We read $x \succ y$ as x is strictly preferred to y or we say that the agent strictly prefers x to y.

Definition: Indifferent Preference Relation $x \sim y$ iff $x \succsim y$ and $y \succsim x$.

We read $x \sim y$ as "x is indifferent to y" or we say that the agent is *indifferent* between x and y.

1.2 Rational Choice

Now lets go back to our definition of *rational choice.* We now have a better understanding of the word *preferred* and this helps us to refine our definition of the term *rational.*

Definition: Rational A preference relation $\succsim$ is *rational* if it possess the following two properties: Rational

1. Completeness
2. Transitivity

Completeness means that if an agent is given a choice between two options, she will have an opinion as to which she likes more. She may be indifferent, but she is never completely clueless. Completeness

Definition: Completeness A preference relation $\succsim$ on X is *complete* if $\forall x, y \in X$ either $x \succsim y$ or $y \succsim x$, or both.

Because this definition does not excludes the possibility that $y = x$, completeness implies that $x \succsim x$.

Transitivity means that if an agent prefers (meaning weakly prefers) option 1 to option 2 and also prefers option 2 to option 3, then she will prefer option 1 to option 3. Transitivity

Definition: Transitivity A preference relation $\succsim$ on X is *transitive* if $\forall x, y, z \in X$ if $x \succsim y$ and $y \succsim z$, then $x \succsim z$.

Transitivity also means that an agent's weak preferences can cycle only among choices that are indifferent. That is, if she weakly prefers A to B, B to C, and C to A, then she must be indifferent among all three: A∼B∼C.

The word *rational* lends a sense of credibility, but remember that rational choice is an assumption. An economist using the standard framework assumes that an individual's preference relation is rational. There are many real-life examples of violations of transitivity. The following example is due to Kahneman and Tversky (1984).

EXAMPLE: Imagine that you are about to purchase a stereo for \$125 dollars and a calculator for \$15. The salesman tells you that the calculator is on sale for 5 dollars less at the other branch of the store, located 20 minutes away. The stereo is the same price there. Would you make the trip to the other store?

The fraction of people saying that they would travel to the other store for the \$5 discount is much higher when the calculator is discounted than when the question is changed so that the \$5 discount applies to the stereo instead. This is so even though the ultimate saving obtained by incurring the inconvenience of travel is the same in both cases. In fact, if asked the following question, we would expect people to respond with indifference:

Both the calculator and the stereo are out of stock, so you must travel to the other store (located 20 minutes away). At the other store, you will receive a \$5 discount coupon as compensation. Do you care on which item we apply this 5 dollar rebate?

However, responding to this question with indifference exposes an intransitivity. Let x be traveling to the other store to get a calculator discount, let y be traveling to get a stereo discount, and let z be staying at the first store. That a high fraction of people respond that they would go to the other store for the calculator discount means that $x \succ z$. That a low fraction of people respond that they would go to the other store for the stereo discount means that $z \succ y$. However, answering the above question with indifference means that $x \sim y$, a clear violation of transitivity.

1.3 Choice Structure

Choice Structure

Given preferences, how will an economic agent behave? We assume that given a set of choices $B \subset X$, the agent will choose the element of B she prefers most. To formalize this, we define the agent's *choice structure* $(\mathfrak{B}, C(.))$.

- $\mathfrak{B}$ is a set of nonempty subsets of X; that is, every element of $\mathfrak{B}$ is a set $B \in X$. We call the elements $B \in \mathfrak{B}$ *budget sets.*

- $C(.)$ is a *choice rule* (actually, it's a correspondence) that assigns a set of chosen elements of B. Formally, $C(B; \succsim) = \{x \in B \mid x \succsim y \; \forall y \in B\}$. The choice rule gives the set of items in B that the agent likes as much as any of the other alternatives.

Choice Rules

There are several things to note about $C(B; \succsim)$.

- $C(B; \succsim)$ may contain more than one element.

- If B is finite, then $C(B; \succsim)$ is non-empty (we will prove this in proposition 1).

- If B is infinite, then $C(B; \succsim)$ might be empty. To see why, suppose $B = \{x \mid x \in [0, 1)\}$. If the agent feels that more is better (so $x \succsim y$ if $x \geq y$), then $C(B; \succsim) = \emptyset$.

An example of how we use this choice structure may help.

> **Example:** Suppose that $X = \{x, y, z\}$ and $\mathfrak{B} = \{\{x, y\}, \{x, y, z\}\}$. One possible choice structure is $(\mathfrak{B}, C_1(.))$, where the choice rule $C_1(\{x, y\}, \succsim) = \{x\}$ and $C_1(\{x, y, z\}, \succsim) = \{x\}$. In this case, x is chosen no matter which decision the individual faces.

If an agent's preferences are complete and transitive, then her choice rule will not be completely arbitrary, as the following result shows.

Proposition 1: *Suppose $\succsim$ is complete and transitive. Then,*

1. *for every finite non-empty set B, $C(B;\succsim) \neq \emptyset$ and*
2. *for A, $B \in \mathfrak{B}$ if $x, y \in A \cap B$, and $x \in C(A;\succsim)$ and $y \in C(B;\succsim)$, then $x \in C(B;\succsim)$ and $y \in C(A;\succsim)$*

Proof: For (1), we proceed by mathematical induction on the number of elements of B. First, suppose the number of elements is one, so $B = \{x\}$. By completeness, $x \succsim x$, so $x \in C(B;\succsim)$. Hence, for all sets B with just one element, $C(B;\succsim) \neq \varnothing$. Next, fix $n \geq 1$ and suppose that for all sets B with exactly n elements, $C(B;\succsim) \neq \varnothing$. Let A be a set with exactly $n+1$ elements and let $x \in A$. Then, there is a set B with exactly n elements such that $A = B \cup \{x\}$. By the induction hypothesis, $C(B;\succsim) \neq \varnothing$, so let $y \in C(B;\succsim)$. If $y \succsim x$, then by definition $y \in C(A,\succsim)$, so $C(A,\succsim) \neq \varnothing$. By completeness, the only other possibility is that $x \succsim y$. In that case, for all $z \in B$, $x \succsim y \succsim z$, so transitivity implies that $x \succsim z$. Since $x \succsim x$, it follows that $x \in C(A;\succsim)$ and hence that $C(A,\succsim) \neq \varnothing$. Hence, for every set A with exactly $n+1$ elements, $C(A,\succsim) \neq \varnothing$. By the principle of mathematical induction, it follows that for every finite set A with any number of elements, $C(A,\succsim) \neq \varnothing$, which proves (1) $\square$

Proof: For (2), if $x, y \in A$, and $x \in C(A;\succsim)$, then $x \succsim y$. The condition $y \in C(B;\succsim)$ means that for all $z \in B$, $y \succeq z$. Then, by transitivity, for all $z \in B$, $x \succsim z$. From that and $x \in B$, we conclude that $x \in C(B;\succsim)$. A symmetric argument implies that $y \in C(A;\succsim)$. $\square$

Q.E.D.

Chapter 2

Choice and Revealed Preferences

While economic theories tend to begin by making assumptions about people's preferences and then asking what will happen, it is interesting to turn this process around. Indeed, much empirical work reasons in the reverse way: it looks at people's choices (e.g. how much money they've saved, what car they bought), and tries to "rationalize" those choices, that is, figure out whether the choices are compatible with optimization and, if so, what the choices imply about the agent's preferences.

Rationalize Choices

What are the implications of optimization?

Question: Can we *always* rationalize choices as being the result of preference maximization? Or does the model of preference maximization have *testable restrictions* that can be violated by observed choices?

2.1 Choice Rule

In the preceding chapter, we derived a choice rule from a given preference relation, writing $C(B; \succsim)$ to emphasize the derivation. In empirical data,

Choice Rule

however, the evidence comes in the form of choices, so it is helpful to make the choice rule the primitive object of our theory.

Definition: Choice Rule A *choice rule* is a function $C : \mathfrak{B} \rightarrow \mathfrak{B}$ with the property that $\forall\, B \in \mathfrak{B}$, $C\,(B) \subseteq B$.

Question: How can we learn an agent's choice rule?

In principle, we can learn an agent's choice rule by watching her in action, however, we have to see her choose from all subsets of X. Suppose we are able to learn an agent's choice rule. Can we tell if her choice behavior is consistent with her maximizing some underlying preferences?

HARP

Definition: HARP A choice function $C : \mathfrak{B} \rightarrow \mathfrak{B}$ satisfies *Houthaker's Axiom of Revealed Preference* if, whenever $x, y \in A \cap B$, and $x \in C(A)$, and $y \in C(B)$, then $x \in C(B)$ and $y \in C(A)$.

Proposition 2: *Suppose* $C : \mathfrak{B} \rightarrow \mathfrak{B}$ *is non-empty. Then* $\exists$ *a complete and transitive preference relation* $\succsim$ *on* X *such that* $C(.) = C(.; \succsim)$ *iff* C *satisfies HARP.*

That is, C could be the result of an agent maximizing complete and transitive (rational) preferences if and only if C satisfies HARP.

Proof: First, suppose that $C(\cdot) = C(\cdot; \succsim)$ is the result of an agent maximizing complete transitive preferences. From the Proposition 1, we know that C must satisfy HARP.

Conversely, suppose C satisfies HARP. Define the "revealed preference relation" $\succsim_c$ as follows: if for some $A \subset X$, $y \in A$ and $x \in C(A)$, then say that $x \succsim_c y$. We need to show three things, namely, that $\succsim_c$ is complete and transitive and that $C(\cdot) = C(\cdot; \succsim_c)$.

1. For completeness, pick any $x, y \in X$. Because C is non-empty, then either $C(\{x,y\}) = \{x\}$ in which case $x \succsim_c y$, or $C(\{x,y\}) = \{y\}$ in which case $y \succsim_c x$, or $C(\{x,y\}) = \{x,y\}$ in which case $x \succsim_c y$ and $y \succsim_c x$.

2. For transitivity, suppose $x \succsim_c y$ and $y \succsim_c z$, and consider $C(\{x,y,z\})$, which by hypothesis is non-empty. If $y \in C(\{x,y,z\})$, then by HARP, $x \in C(\{x,y,z\})$. If $z \in C(\{x,y,z\})$, then by HARP, $y \in C(\{x,y,z\})$. So, in every possibility, $x \in C(\{x,y,z\})$. Hence, $\succsim_c$ is transitive.

3. If $x \in C(A)$ and $y \in A$, then by the definition of $\succsim_c$, $x \succsim_c y$. So, $x \in C(A; \succsim_c)$. This implies that $C(A) \subset C(A; \succeq_c)$. Also, since $C(A)$ is non-empty, there is some $y \in C(A)$. If $x \in C(A; \succsim_c)$, then $x \succsim_c y$, so by HARP, $x \in C(A)$. This implies that $C(A; \succsim_c) \subset C(A)$. $\square$

Q.E.D.

A problems with Proposition 2 is that it assumes that we observe the entire choice function $C(.)$. Real data is almost never this comprehensive. Typically, if we are trying to test the preference-based model of choice, we will observe less than all of C in two respects:

1. For sets $B \subset X$ where $C(B)$ contains more than one element, we are likely to see only one element of $C(B)$.

2. We will typically see $C(B)$ for some, but not all subsets $B \subset X$. For example, in consumer choice problems, the relevant sets B may be only the budget sets, which is a particular sub-collection of the possible sets B.

To develop a theory based on more limited observations, economists have developed the *weak axiom of revealed preference.* WARP

WARP says that if x is every chosen when y is available, then there can be no budget set containing both alternatives for which y is chosen and x is not.

Definition: WARP The choice structure $(\mathfrak{B}, C(.))$ satisfies the *weak axiom of revealed preference* if the following property holds:

- If for some $B \in \mathfrak{B}$ with $x, y \in B$ we have $x \in C(B)$, then for any $A \in \mathfrak{B}$ with $x, y \in A$ and $y \in C(A)$, then we must also have $x \in C(A)$.

Example: Suppose that $X = \{x, y, z\}$, $\mathfrak{B} = \{\{x, y\}, \{y, z\}, \{x, z\}\}$, $C(\{x, y\}) = \{x\}$, $C(\{y, z\}) = \{y\}$, $C(\{x, z\}) = \{z\}$. This choice structure satisfies the weak axiom. nevertheless, we cannot have rationalizing preferences because transitivity would be violated. Therefore, there can be no rationalizing preference relation.

Question: What condition would we need in addition to the weak axiom being satisfied for there to be a rational preference relation $\succeq$ that rationalizes $C(.)$ relative to $\mathfrak{B}$; that is $C(B) = C(B, \succsim)\ \forall B \in \mathfrak{B}$?

Answer: $\mathfrak{B}$ includes all subsets of X of up to three elements.

For additional material on WARP, see section 1.D of MWG.

Chapter 3

Utility

So far, we have a pretty abstract model of choice. As a step toward having a more tractable mathematical formulation of decision-making, we now introduce the idea of utility, which assigns a numerical ranking to each possible choice. For example, if there are n choices ranked in order from first to last, we may assign the worst choice(s) a utility of 0, the next worst a utility of 1, and so on. Picking the most preferred choice then amounts to picking the choice with the greatest utility. In this chapter, we will examine the classical, preference-based approach to utility functions.

Definition: Utility Function A preference relation $\succsim$ on X is represented by a utility function $u : X \to \mathbb{R}$ if $x \succsim y \Leftrightarrow u(x) \geq u(y)$.

Utility Representation

That is, a utility function assigns a number to each element in X. A utility function u represents a preference relation $\succeq$ if the numerical ranking u gives to elements in X coincides with the preference ranking given by $\succsim$.

Having a utility representation for preferences is convenient because it turns the problem of preference maximization into a relatively familiar math problem. If u represents $\succsim$, then

$$C(B;\succsim) = \left\{ x \mid x \text{ solves } \max_{x\in B} u(x) \right\}.$$

A natural question is whether given a preference relation $\succsim$, we can always find a function u to represent $\succsim$.

Proposition 3: *If X is finite, then any complete and transitive preference relation $\succsim$on X can be represented by a utility function $u : X \to \mathbb{R}$.*

Proof. The proof is by induction on the size of the set. We prove that the representation can be done for a set of size n such that the range $\{u(x)|x \in X\} \subset \{1, ..., n\}$. We begin with $n = 0$. If $X = \varnothing$, then $\{u(x)|x \in X\} = \varnothing$, so the conclusion is trivial. Next, suppose that preferences can be represented as described for any set with at most n elements. Consider a set X with $n + 1$ elements. Since $C(X,\succsim) \neq \varnothing$, the set $X - C(X,\succsim)$ has no more than n elements, so preferences restricted to that set can be represented by a utility function $u$whose range is $\{1, ..., n\}$. We extend the domain of u to X by setting $u(x) = n + 1$ for each $x \in C(X,\succsim)$. Next, we show that this u represents $\succsim$.

Given any $x, y \in X$, suppose that $x \succsim y$. If $x \in C(X,\succsim)$, then $u(x) = n + 1 \geq u(y)$. If $x \notin C(X,\succsim)$, then by transitivity of $\succsim$, it must also be true that $y \notin C(X,\succsim)$, so $x, y \in X - C(X,\succsim)$. Then, by construction, $n + 1 \geq u(x) \geq u(y) \geq 1$.

For the converse, suppose it is not true that $x \succsim y$. Then, $y \succ x$, and a symmetric argument to the one in the preceding paragraph establishes that $n + 1 \geq u(y) > u(x) \geq 1$. Hence, $x \succsim y$ if and only if $u(x) \geq u(y)$, so u represents $\succsim$. □ *Q.E.D.*

If X is infinite, things are a bit more complicated. In general, not every complete, transitive preference relation will be representable by a real-valued utility function. For example, consider the lexicographic preferences according to which $x \succ y$ whenever either

Lexicographic Preferences

1. $x_1 > y_1$ or
2. $x_1 = y_1$ and $x_2 > y_2$.

In words, as long as the first component of x is larger than that of y, x is preferred to y, regardless of the values of the second components of x and y. If the first component s become equal, only the second components are relevant. The name lexicographic is derived from the way a dictionary is organized.

These preferences cannot be represented by a real-valued utility function. The problem is that the agent's preferences are so "refined" that the set of real numbers is too small to capture them all. This is also an example for which indifference curves don't exist, because the agent is never indifferent between any two choices.

3.1 Interpersonal Comparisons

In the first problem set, you were asked to prove that if the preference relation $\succsim$ on X is represented by a utility function $u : X \rightarrow \mathbb{R}$, then it is also represented by $v(u(x))$, for any increasing function $v : \mathbb{R} \rightarrow \mathbb{R}$. This arbitrariness in the way preferences are represented has important consequences.

The intellectual history of the utility idea had its roots in utilitarian theory, according to which, for example, some goods might be more valuable to me than to you, in the sense of giving me more additional utility. For example, suppose giving you some good increases your utility by one, where if the good were given to me, my utility would increases by two, given some representation. If one multiplies all of your utilities by four, one gets an equally valid representation of your preferences, but now the extra utility you get, which is four, exceeds my extra utility of two. This example illustrates the general principle that utilities derived from observed choices cannot be used for "interpersonal comparisons," that is, they cannot definitively resolve questions about the relative value of various goods to you and to me.

Several variants of utility theory have been developed that might, in principle, be used for interpersonal comparisons. These variants give meaning to statements like "food means more to a starving person than to a sated person."

One famous variant is based on choices made *behind the veil of ignorance*, in which individuals are asked to consider the possibility that they might have become either the starving person or the sated person. Behind the philosophically motivated veil of ignorance, they do not know which position they will occupy and are asked to decide, from that perspective, what rule they would hope applies to allocate any extra food that might become available. This attempt to base a moral decision on a choice behind the veil of ignorance will be discussed again after we have treated the theory of choice under uncertainty. A criticism of this entire approach is that it is based on hypothetical choices, rather than real ones. Hypothetical choices, critics argue, do not have the same standing as real choices and are not a reliable way to predict real choices.

Another variant develops a different conception of utility theory, based on the idea that people do not notice small differences and express indifference between choices that are close. If that accurately describes human behavior, then one might determine a unit for measuring utility by determining empirically the *just noticeable difference.* The idea is to derive a utility representation in which $x \succ y \iff u(x) \geq u(y) + 1$ and so $x \sim y \iff not(x \succ y$ or $y \succ x) \iff |u(x) - u(y)| < 1$. This is a theory in which $\succ$ is transitive but $\sim$ is not, so it is fundamentally different from the theory described above. This theory formalizes the idea that a starving person may benefit more from extra food than a sated person. In fact, if a small amount of food were transfered from the sated person, she might not even notice that it was missing while the hungry person would enjoy a clear benefit.

These and other theories of choice are omitted here so that we can focus our limited attention on the details of rational choice, which is the model that underlies the overwhelming majority of economic analysis. Ironically, however, the very acknowledgement that we do have limited attention that can affect our choice of what to study is itself a challenge to the classic model of rational choice.

3.2 Restrictions on Preferences

To make progress in economics research, it is almost always necessary to make additional assumptions that restrict preferences in various ways. Economists try to be careful about these assumptions. We make the minimal assumption necessary for the analysis to be tractable and investigate all the implications of any assumption, so that they can be tested using whatever data is available.

In this section, we will look at several restrictions on preferences that are the most commonly used ones in economic analysis. Our task in each case is to identify how restrictions on preferences and restrictions on utility functions are related. On one hand, since modelers usually work with utility functions, the idea is sometimes to identify all the restrictions on choices implied by a particular assumption about utility. On the other hand, when the desired restriction on choices is given, the problem is to identify the exact restriction on utility functions that characterize the given restriction on choices.

3.2.1 Continuity

Continuity

The following continuity restriction on preferences is a condition that implies not only that a utility representation exists, but that a continuous representation exists. This restriction is usually considered uncontroversial in empirical science for the simple and compelling reason that any *finite* set of observed choices that is consistent with HARP is also consistent with continuity. That is, if the data take the form of observations of choices, then continuity can be contradicted only by an *infinite* data set.

Definition: Continuous A preference relation $\succsim$ on X is continuous if for any sequence $\{(x^n, y^n)\}_{n=1}^{\infty}$ with $x^n \to x$, $y^n \to y$, and $x^n \succsim y^n \, \forall \, n$, we have that $x \succsim y$.

Continuity says that the consumer's preferences cannot exhibit jumps. For example, an individual with a continuous preference relation cannot prefer each element in sequence $\{x^n\}$ to the corresponding element in sequence $\{y^n\}$ but then suddenly reverse her preferences at the limiting points of the sequences, x and y.

An equivalent definition of continuity is the following: $\forall \, x$, the upper contour set $\{y \in X \mid y \succsim x\}$ and the lower contour set $\{y \in X \mid x \succsim y\}$ are both closed; that is, they include their boundaries.

For an example of preferences that are not continuous, we can again use the lexicographic preferences described on page 15. Consider the sequence of bundles $x^n = (\frac{1}{n}, 0)$ and $y^n = (0, 1)$. For every n, $x^n \succsim y^n$. However, $\lim y^n = (0, 1) \succsim (0, 0) = \lim x^n$.

Continuity of a rational preference relation guarantees the existence of a continuous utility function representation, as stated in Proposition 4.

Proposition 4: *If $X \subset \mathbb{R}^n$, then any complete, transitive, and continuous preference relation $\succeq$ on X can be represented by a continuous utility function $u : X \to \mathbb{R}$.*

Proof: We prove this for the case of a strictly monotone preference relation $\succsim$, and $X = \mathbb{R}^n_+$. (Note: A preference relation $\succsim$ is strictly monotone if $x >> y$ implies that $x \succ y$, that is, if the consumer always prefers a bundle that provides more of every good.) This extra restriction allows a simple, constructive proof.

Let $e = (1, ..., 1)$ denote one bundle in x, and consider elements of $\mathbb{R}^n_+$of the form $\alpha e = (\alpha, ..., \alpha)$ where $\alpha \geq 0$. We claim that for any $x \in \mathbb{R}^n_+$, there exists a *unique* value $\alpha(x) \in [0, \infty)$ such that $\alpha(x) \sim x$. We will then construct u by letting $u(x) = \alpha(x)$.

To prove the claim, let $x \in \mathbb{R}^n_+$ be given. Consider the two sets

$$\begin{aligned} A^+ &= \{\alpha \in \mathbb{R}_+ : \alpha e \succsim x\}, \\ A^- &= \{\alpha \in \mathbb{R}_+ : x \succsim \alpha e\}. \end{aligned}$$

Both sets are non-empty. To see why, note that $x \succsim 0$ by monotonicity, so A^- in non-empty, and if we choose $\overline{\alpha}(x)$ such that $\overline{\alpha}(x)e \gg x$, then by monotonicity $\overline{\alpha}e \succsim x$ and $\overline{\alpha} \in A^+$. By the continuity of preferences, both sets are also closed, and their union is $\mathbb{R}_+$, so $A^+ \cap A^-$ is not empty. Also, for any $\alpha' > \alpha$, monotonicity implies that $\alpha' e \succ \alpha e$, so $A^+ \cap A^-$ contains at most one element: call it $\alpha(x)$.

Now for every $x \in \mathbb{R}^n_+$, we specify the utility by $u(x) = \alpha(x)$. We need to show that this utility function (1) represents the preference relation $\succsim$, and (2) is continuous. For the representation part, suppose that $\alpha(x) \geq \alpha(y)$. Using monotonicity, $x \sim \alpha(x)e \succsim \alpha(y)e \sim y$, so $x \succsim y$. Conversely if $x \succsim y$, then by construction $\alpha(x)e \sim x \succsim y \sim \alpha(y)e$, and by monotonicity

$\alpha(x) \geq \alpha(y)$. Continuity is more subtle and we omit it (see MWG, Section 3.C if you're interested in a proof). □

Q.E.D.

Intuitively, the construction in the proof specifies the utility of any choice x by finding the point at which its indifference surface crosses the 45° line. This specification is, of course, completely arbitrary. It is simply a mathematically convenient way to represent someone's preferences.

3.2.2 Monotonicity and Local Non-Satiation

The next restrictions examined below are *monotonicity* and *local non-satiation*, which are used extensively in Part II. Roughly put, these imply that consumers will prefer to spend all of their wealth or income on something, because more is always at least as good as less and consumers are never satiated. This conclusion about consumer spending will be a useful intermediate step for making inferences from consumer's observed choices.

Monotonicity

DEFINITION: Monotonicity A preference relation $\succsim$ on X is monotone if for any x, $y \in X$, $\{x_i \geq y_i\}_{i=1}^n$ implies that $x \succsim y$.

Monotonicity of preferences makes sense if X represents bundles of goods, so that if $x = (x_1, .., x_n)$, then x_k is the amount of good k. If "more of a good is good," then preferences will be monotone.

Local Non-Satiation

DEFINITION: Local Non-Satiation A preference relation $\succsim$ on X is locally non-satiated if for any $y \in X$, and $\epsilon > 0$, $\exists\ x \in X \cap B_\epsilon(y)$ such that $x \succ y$.

If an agent's preferences are locally non-satiated, this means that there is no bundle in X that is "ideal" or even locally ideal. There is always a slight change that would leave the agent better off. Notice that local non-satiation is a joint property of the preference relation $\succsim$ and the choice set X. If the choice set involves only integer quantities of *every* good, then the preferences cannot be locally non-satiated. However, the condition does not require that all goods be divisible; it is satisfied, for example, in every model for which one good (say, leisure time) is always valuable and divisible.

3.2.3 Convexity

The next restriction is that consumer preferences are *convex*. Convexity is used to infer the existence of prices for various purposes. It is fundamental for studying markets because in the standard model of competitive economies, when consumer preferences are convex, market clearing prices exist. In general, when preferences are not convex, they may not exist. In the latter case, one cannot sustain the common hypothesis that a competitive, market-clearing outcome approximates actual market outcomes. Convexity is also used in discussions about whether consumer preferences are recoverable from choices from various budget sets. In that application, convexity is used to establish that there are prices which cause different preferences to lead to different choices.

Convex Preferences

Definition: Convex A preference relation $\succsim$ on a convex choice set X is convex if $x \succsim y$ and $x' \succsim y$ implies that for any $t \in [0, 1]$, $tx + (1-t)x' \succsim y$.

Convexity is often described as capturing the idea that agents like diversity. That is, if an agent is indifferent between x and y and has convex preferences, then she will like $\frac{1}{2}x + \frac{1}{2}y$ at least as much as either x or y. Of course, this doesn't always make sense. You might like a root beer or orange juice, but not a mixture.

Whether convexity makes sense often depends on the interpretation of the goods space. For example, if the components of x are rates of consumption, then a half-half mixture of root beer and orange juice might mean drinking root beer half the time and orange juice half the time. Convexity of preferences seems more plausible in that interpretation than in the previous one. Similarly, some find convexity easier to rationalize if the goods are more highly aggregated — for instance, if the goods are food and clothing, than if goods are highly specific.

An equivalent way to describe convexity uses the indifference curves and surfaces of undergraduate economics: convexity of preferences amounts to the assumption that the upper contour set of any $y \in X$–meaning the set of points above the indifference surface through y–is a convex set. Formally, the upper contour set is given by

$$\text{Upper Contour Set of } y = \{x \in X : x \succsim y\}.$$

There is also a related notion of *strict convexity,* which says that if $x \succsim y$ and $x' \succsim y$ and $x \neq x'$ then for any $t \in (0, 1)$, $tx + (1 - t)x' \succ y$.

Figure 3.1 is a representation of preferences that are not convex. Figure 3.2 is a representation of preferences that are convex but not strictly convex. Figure 3.3 is a representation of preferences that are strictly convex.

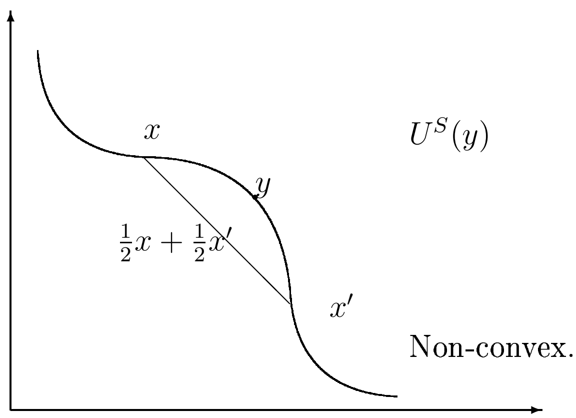

Figure 3.1: Non-Convex Preferences

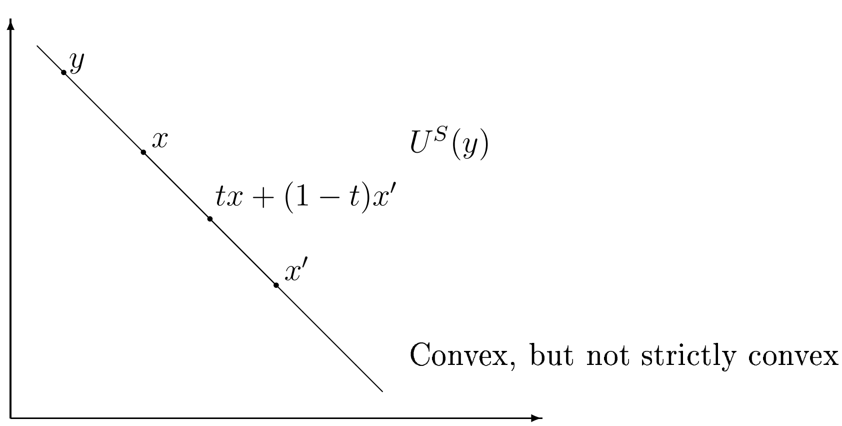

Figure 3.2: Convex Preferences.

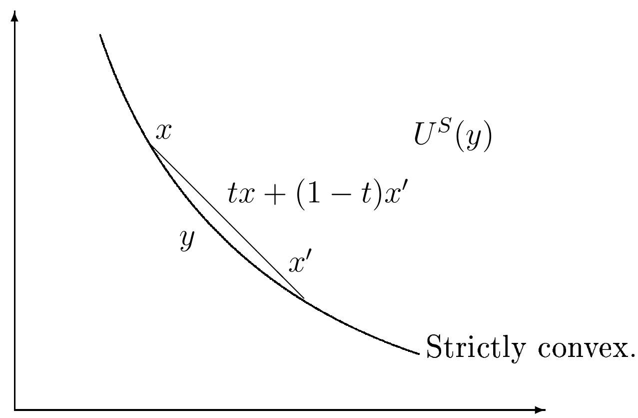

Figure 3.3: Strictly Convex Preferences.

Each of these properties of preferences has a corresponding property if we look at a preference-representing utility function.

Proposition 5: *Suppose the preference relation $\succsim$ on X can be represented by $u : X \to R$. Then:*

1. $\succsim$ *is monotone iff u is nondecreasing*
2. $\succsim$ *is locally non-satiated iff u has no local maxima in X*
3. $\succsim$ *is (strictly) convex iff u is (strictly) quasi-concave.*

Quasi Concavity

Note that the utility function $u(.)$ is *quasi-concave* if the set $\{y \in \mathbb{R}^n_+ \mid u(y) \geq u(x)\}$ is convex $\forall x$, or, equivalently, if $u(\alpha x + (1-\alpha)y) \geq \min\{u(x), u(y)\}$ for any x, y and all $\alpha \in [0,1]$.

3.2.4 Separability

Separability

A common assumption made by empirical researchers into consumer behavior is *separability*.

For example, a researcher might postulate that a consumer's decision about how to divide her total entertainment spending among various kinds of entertainment options (such as movies, concerts, clubs, and so on) does not depend on her choices about housing, food, clothing, etc. When this assumption can be validated, it simplifies empirical work because, for example, it implies that the researcher can legitimately estimate the demand for entertainment goods based on (1) total entertainment spending and (2) the prices of entertainment goods, even without information about the prices of food, housing and clothing.

Suppose a consumer is choosing a bundle of goods in $\mathbb{R}^n$. There may be some number of goods $m < n$ that the consumer regards as a natural group. For example, they may be different kinds of entertainment goods described by the vector $x \in \mathbb{R}^m$, while the other goods are described by the vector $y \in \mathbb{R}^{n-m}$, so that the consumer's overall choice is described by $(x, y) \in \mathbb{R}^n$. We investigate when it is possible to identify the consumer's vector of entertainment choices of x from limited information, namely, information about total entertainment spending and the prices of various entertainment goods, without any further information about y or the consumer's total spending. When that is possible, we say that the choice of x does not depend on y.

> **EXAMPLE:** Suppose that the choices x are various kinds of entertainment, while the choices y include restaurant meals, home meals, and housing. Suppose that the decision to purchase restaurant meals is closely related to entertainment, for example because one eats out more often when attending the movie or a concert or, reversely, because a leisurely dinner out is a substitute for other entertainment. In that case, the overall level of entertainment spending could affect the choice between home meals and restaurant meals, even if the overall level of food spending

doesn't affect the choice between movies and concerts.

Notice, too, that separability can be layered in various ways. A utility function might have the form $u(x, y) = U(v_x(x), v_y(y))$, which gives symmetric separability. Another possibility is that $u(x, y, z) = U(V(v(x), y), z)$, where V and v are real-valued functions and V and U are each increasing in the first argument. This would imply both that the choice of x does not depend on (y, z).and that the choice of (x, y) does not depend on z. This might represent the preferences of a consumer who regards restaurant meals y as in the entertainment category, separable from non-entertainment decisions, and also regards the choice x between attending movies or concerts as independent of the quantity of restaurant meals.

We have just scratched the surface of separability. For example, economists sometimes wish to create a *price index* for entertainment goods that depends only on the prices of those goods. Such an index is useful if, together with the other prices and the consumer's income, it determines the total spending on entertainment goods and the consumer's welfare. The question of whether it is possible to create a price index for a category of goods, independent of the prices of other goods, is logically distinct from the question studied above. The index sought here is used to characterize how much is *spent in total on entertainment goods* whereas above we asked about how any total spending would be allocated *among entertainment goods.*

3.2.5 No Wealth Effects

Wealth Effects

Finally, much current research about organizations and transactions costs relies on the idea that there are *no wealth effects*. This means that consumer choices about how much to buy of certain goods are unaffected by wealth transfers (at least within certain ranges), so efficient allocations can be determined without knowing how wealth is distributed. This is a very special situation, and the associated question is the usual one: what must the utility function look like for this no-wealth-effects property to be satisfied?

Let us suppose that $X = \mathbb{R}_+ \times Y$, which we interpret to means that the choice space consists of a quantity of some one good and some other choices.

Proposition 6: *Suppose the preference relation $\succsim$ on $X = \mathbb{R}_+ \times Y$ is complete and transitive and that there exists $\bar{y} \in Y$ such that $\forall\, y \in Y$, $(0,\, y) \succsim (0,\bar{}\, y)$. If*

1. *"Good 1 is Valuable"* $(a,\, \bar{y}) \succsim (a',\, \bar{y})$ *iff* $a \geq a'$.
2. *"Compensation is Possible" for every $y \in Y$, there exists a $t \geq 0$ such that $(0,\, y) \sim (t,\, \bar{y})$.*
3. *"No Wealth Effects" if $(a,\, y) \succsim (a',\, y')$ then $\forall\, t \in \mathbb{R}$, $(a + t,\, y) \succsim (a' + t,\, y)$.*

Then $\exists\; v : Y \to \mathbb{R}$ such that $(a,\, y) \succsim (a',\, y')$ iff $a + v(y) \geq a' + v(y')$. Conversely, if the preference relation $\succsim$ on $X = \mathbb{R} \times Y$ is represented by $u(a,\, y) = a + v(y)$, then it satisfies the three preceding conditions.

Proof: By the second condition, we may define a function $v(y)$ so that for each $y \in Y$, $(0,\, y) \backsim (v(y),\, \overline{y})$. By the third condition, for any $(a,\, y)$, $(a',\, y') \in \mathbb{R} \times Y$, $(a,\, y) \backsim (a + v(y),\, \overline{y})$ and $(a',\, y') \backsim (a' + v(y'),\, \overline{y})$. So by transitivity, $(a,\, y) \succsim (a',\, y')$ if and only if $(a + v(y),\, \overline{y}) \succsim (a' + v(y'),\, \overline{y})$. By the first condition, that is equivalent to $a + v(y) \geq a' + v(y')$. Hence, the three conditions imply that $(a,\, y) \succsim (a',\, y')$ if and only if $a + v(y) \geq a' + v(y')$.

It is routine to verify the converse, namely, that the representation implies the three conditions. □

Q.E.D.

The first condition of the proposition implies the local non-satiation condition. The second condition ensures that good one is sufficiently valuable that some amount of it will compensate for any change in y. The condition is most reasonable for applications in which all the relevant goods are traded in markets. It is certainly possible to imagine choice problems in which compensation of this sort is not possible. A person's choices might reflect a conviction that there is no way for cash or market transactions to compensate fully for poor health or for the loss of one's child, or for an increased likelihood of going to heaven.

The third condition is a subtle one, combining elements of separability and framing. It asserts separability, because it asserts that the choice between two alternatives does not depend on the consumer's initial endowment of good 1. In addition, because the objects of choice are outcomes and transfers, it allows the choice to be *framed* in terms of *changes* in good 1, rather than the level of consumption of good 1.

The form $a + v(y)$ is called the *quasi-linear* form and is usually used with the good 1 interpreted as "money" or the *numeraire* good. Particularly in the theory of the firm, a profit-maximizing firm is often assumed to convert all outcomes of any sort into a money equivalent that it adds to its cash profits as a criterion to evaluate complex outcomes.

3.3 A Few Criticism of Rational Choice

The central position of the rational choice model in economic analysis does not mean that it is beyond challenge. On the contrary, its very centrality means that it is important for economists to be aware of its role and limits. There is a long tradition of research marshaling experimental and empirical evidence that is in conflict with the most basic rational choice model. Recently, there is a growing movement that questions the model's assumptions and seeks to incorporate insights from psychology, sociology and cognitive neuroscience into economic analysis.

Preference Foundation

What does it mean for preferences to be the *foundation* for this rational choice approach to economic modeling? Suppose for example we wanted to compare the relative merits of capitalism and socialism. The rational choice approach would start by specifying the relevant preferences (for example, everyone likes to consume more, some people might not like inequality, and so on), model the allocation of resources under capitalism and socialism and then make a comparison as to which system people prefer by comparing the outcomes of the systems. This approach *takes preferences as primitives*, abstracting from the idea that preferences themselves are affected by institutions — this is the sense in which preferences are foundational. Without that, the comparative welfare analysis of the different systems would be much trickier.

Usually (but not always), rational choice approaches are also *consequentialist;* they abstract also from the logical possibility that people care about process, rather than merely economic outcomes. Of course, institutions clearly do affect preferences and some people are willing to exchange worse economic outcomes for a sense of control. To some extent, the standard model can be relaxed to accommodate these points, but generally speaking it is precisely these simplifications — that preferences are fundamental, focused on outcomes, and not too easily influenced by one's environment — that allows economic analysis to yield sharp answers to some interesting public policy questions.

A main criticism of the most basic rational choice model is that real-world choices often appear to be highly situational or context-dependent (meaning that an individual's preferences are easily influenced by her environment). The way in which a choice is posed, the social context of the decision, the emotional state of the decision-maker, the addition of seemingly extraneous items to the choice set, and a host of other environmental factors appear to influence choice behavior. The existence of the marketing industry is testament to this, and many other examples are possible. A simple example is the presence of a tempting chocolate cake on the dessert menu might make you feel good about sharing an order of apple pie, when you might have ordered fruit instead if you hadn't been tempted by the chocolate cake.

Strictly speaking, there is little formal problem in allowing preferences to depend on context. That said, the strength of the rational choice model derives from the assumption that preferences are relatively stable and not too situation-dependent. This is the source of the theory's empirical content, because it allows us to observe choices in one situation and then draw inferences about choices in related situations. Such inferences become problematic if preferences are highly sensitive to context.

Cognitive Limits

A strong assumption of the basic choice model is that agents have essentially unlimited cognitive capacity. If they are faced with a choice set many times, they will always solve the preference optimization problem correctly. In the standard model, there is no randomness or mistakes in the agent's behavior. A criticism is that in reality, many choices are not considered. Rather they are based on intuitive reasoning, heuristics or instinct. That people rely on intuition and heuristics is not surprising. Given that people have limited cognitive capacity, there is simply no way to reason through every decision. Arguably, instinctive judgement may often mimic preference maximization, particularly in familiar environments. When people rely on heuristic reasoning or intuition in unfamiliar situations, however, the results can be striking departures from the sort of behavior predicted by rational choice models.

Particularly surprising behavior can result when people in unfamiliar situations are given inappropriate contextual clues. For instance, Ariely, Loewenstein and Prelec (2003, QJE) report an experiment in which they showed students in an MBA class six ordinary products (wine, chocolate, books, computer accessories). The items had an average retail price of about $70. Students were asked whether they would buy each good at an amount equal to the last two digits of their social security number. They were then asked to state their valuation for each good.

In spite of the familiarity of the products, students' reported valuations correlated significantly with the random final digits of their social security number. That is, it appears that the students had no firm valuation in mind and "anchored" their value to an essentially arbitrary suggestion (the final two digits of the social security number). Interestingly, Ariely, Loewenstein and Prelec go on to show that once people have fixed on a valuation, they respond to price changes, and other changes in ways that are consistent with the rational choice model. The authors label this behavior "coherent arbitrariness."

A second example that has attracted much attention is the role of default choices. For instance, Madrian and Shea (2001, QJE) provide evidence that enrollment in employer-sponsored 401-K retirement plans (an extremely good deal for most workers by objective criteria) is highly sensitive to whether workers must "opt-in" or "opt-out" of the plan. Another example along these lines comes from organ donations. In the United States, people must opt-in to become a donor by signing up when they get their driver's license. There is a dire shortage of organ donors relative to needy recipients. In Spain, people must opt-out and there the supply greatly exceeds the demand.

The behavior in these examples is hard to square happily with the most basic preference maximization approach. Once one tries to move away from optimization, however, modeling becomes a difficult challenge. That being said, there are models of decision-making that acknowledge people's limited cognitive capacity. These models take a variety of forms: some assume that people make systematic "mistakes" or optimize only partially; others assume people used mixed learning rules, or rules of thumb. It is safe to say however, that there is plenty of work left to be done in developing better models that incorporate bounded rationality.

Part II

Consumer Theory

Chapter 4

Consumer Choice

Since Marshall, the standard approach to developing a theory of competitive markets is to separate demand behavior (consumer theory) from supply behavior (producer theory) and then use the notion of market equilibrium to reconcile demand and supply. Here we will study consumer theory, in Part III, we will study producer theory, and in Part V, we will explore what is meant by a market equilibrium.

4.1 The Consumer Problem

Consumer theory is concerned with how a rational consumer would make consumption decisions. Under assumptions that we explored in our notes on choice theory, a consumer's preferences over the infinite set of consumption bundles can be represented using a continuous utility function. What makes the consumer problem worthy of separate study, apart from the general problem of choice theory, is that the sets from which a consumer chooses are assumed to be defined by certain prices, p, and by the consumer's income or wealth, w. With that motivation in mind, we define the *consumer problem* (CP) to be:

Consumer Problem

$$\max_{x \in \mathbb{R}^n_+} u(x)$$
$$\text{such that } p \cdot x \leq w$$

The idea is that the consumer chooses a vector of goods $x = (x_1, ..., x_n)$ to maximize her utility subject to a *budget constraint* $p \cdot x \leq w$ that says she cannot spend more than her total wealth.

Budget Constraint

What exactly is a "good"? The answer lies in the eye of the modeler. Depending on the problem to be analyzed, goods might be very specific, like tickets to different world series games, or very aggregated like food and shelter, or consumption and leisure. The components of x might refer to quantities of different goods, as if all consumption takes place at a moment in time, or they might refer to average rates of consumption of each good over time. If we want to emphasize the roles of quality, time and place, the description of a good could be something like "Number 2 grade Red Winter Wheat in Chicago." Of course, the way we specify goods can affect the kinds of assumptions that make sense in a model. Some assumptions implicit in this formulation will be discussed below.

Budget Set

Given prices p and wealth w, we can write the agent's *budget set*:

$$B(p, w) = \left\{x \in \mathbb{R}^n_+ \mid p \cdot x \leq w\right\}$$

The consumer's problem is to choose the element $x \in B(p, w)$ that is most preferred or, equivalently, that has the greatest utility. If we restrict ourselves to just two goods, the budget set has a nice graphical representation, as is shown in Figure 4.1.

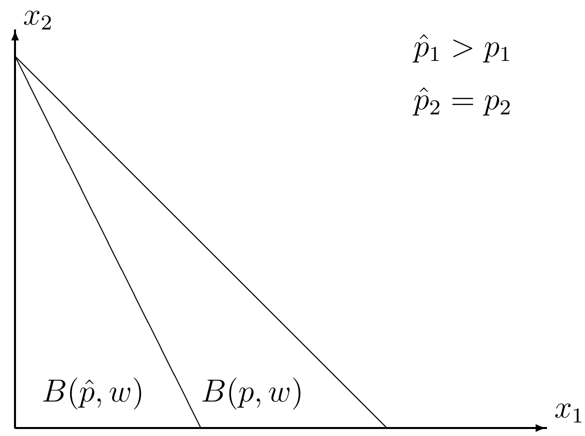

Figure 4.1: The budget set at different prices.

Let's make a few observations about the model:

1. The assumption of *perfect information is* built deeply into the formulation of this choice problem, just as it is in the underlying choice theory. Some alternative models treat the consumer as rational but uncertain about the products, for example how a particular food will taste or a how well a cleaning product will perform. Some goods may be *experience goods* which the consumer can best learn about by trying ("experiencing") the good. In that case, the consumer might want to buy some now and decide later whether to buy more. That situation would need a different formulation. Similarly, if the agent thinks that high price goods are more likely to perform in a satisfactory way, that, too, would suggest quite a different formulation. Perfect Information

2. Agents are *price-takers*. The agent takes prices p as known, fixed and exogenous. This assumption excludes things like searching for better prices or bargaining for a discount. Price Taking

3. Prices are *linear*. Every unit of a particular good k comes at the same price p_k. So, for instance, there are no quantity discounts (though Linear Prices

these could be accommodated with relatively minor changes in the formulation).

Divisibility

4. Goods are *divisible*. Formally, this is expressed by the condition $x \in \mathbb{R}^n_+$, which means that the agent may purchase good k in any amount she can afford (e.g. 7.5 units or π units). *Notice that this divisibility assumption, by itself, does not prevent us from applying the model to situations with discrete, indivisible goods.* For example, if the commodity space includes automobile of which consumers may buy only an integer number, we can accommodate that by specifying that the consumer's utility depends only on the integer part of the number of automobiles purchased. In these notes, with the exception of the theorems that assume convex preferences, all of the results remain true even when some of the goods may be indivisible.

4.2 Some Topological Preliminaries

Before we can continue studying consumer choice, we need two additional definitions.

DEFINITION: Closed A set, S, is closed iff the complement of S, $\backslash S$, is open. Alternatively, S is closed iff all convergent sequences in S have limits in S.

Question: Let $S = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$. Is S closed?

Answer: Yes, the complement of S is an open set, because for $(x, y) \in \backslash S$, the epsilon ball B_ϵ of radius $\epsilon = \sqrt{x^2 + y^2} - 1$ will be entirely contained in $\backslash S$.

DEFINITION: Bounded A set, $S \subset \mathbb{R}^n$, is bounded iff $\exists\ M \in \mathbb{R}_+$ such that all $s \in S$ are contained within the M-ball, B_M.

Chapter 5

Marshallian Demand

In this section and the next, we derive some key properties of the consumer problem.

5.1 Properties of Budget Sets

PROPOSITION 1: (Budget Sets) *Budget sets are homogeneous of degree 0: for all $\lambda > 0$, $B(\lambda p, \lambda w) = B(p, w)$.*

Proof. For $\lambda > 0$, $B(\lambda p, \lambda w) = \{x \in \mathbb{R}^n_+ | \lambda p \cdot x \leq \lambda w\} = \{x \in \mathbb{R}^n_+ | p \cdot x \leq w\} = B(p, w)$. $\square$

Q.E.D.

PROPOSITION 2: (Compact) *If $\{p_i > 0\}_{i=1}^n$, then $B(p, w)$ is compact.*

Proof. If $\{p_i > 0\}_{i=1}^n$, then $B(p, w)$ is a closed and bounded subset of $\mathbb{R}^n_+$. Hence, by the Heine-Borel theorem[1], it is compact. □

Q.E.D.

5.2 Solution to the Consumer Problem

PROPOSITION 3: (Existence) *If u is a continuous function and $\{p_i > 0\}_{i=1}^n$, then the consumer problem (CP) has a solution.*

Proof. By the maximum-minimum theorem, a continuous function on a compact set achieves its maximum. □

Q.E.D.

Question: Does the function $f(x) = x$ on the interval $[0, 1)$ achieve its maximum?

Answer: No, even though there are an infinite number of points as near to 1 as we please, there is no point x for which $f(x) = 1$. However, if the set were compact, for example the set $[0, 1]$, then any continuous function on that set achieves its maximum.

Marshallian Demand

We call the solution to the consumer problem, $x(p, w)$, the *Marshallian demand (or Walrasian demand or uncompensated demand)*. In general, $x(p, w)$ is a set, rather than a single point. Thus $x : \mathbb{R}^n_+ \times \mathbb{R}_+ \rightrightarrows \mathbb{R}^n_+$ is a *correspondence*. It maps prices $p \in \mathbb{R}^n_+$ and wealth $w \in \mathbb{R}_+$ into a *set* of possible consumption bundles. One needs more assumptions (we're getting there) to ensure that $x(p, w)$ is single-valued, so that $x(\cdot, \cdot)$ is a function.

[1] The Heine-Borel theorem states that a set in $\mathbb{R}^n$ is compact iff it is closed and bounded. This result is special to the metric space $\mathbb{R}^n$. A compact set is closed and bounded in any metric space; note however that a closed and bounded set is not necessarily compact, unless it is a subset of $\mathbb{R}^n$.

5.3 Properties of Marshallian Demand

PROPOSITION 4: (Homogeneity) *Marshallian demand is homogeneous of degree zero: for all* p, w, *and* $\lambda > 0$, $x(\lambda p, \lambda w) = x(p, w)$.

Proof. Since by proposition 1, $B(\lambda p, \lambda w) = B(p, w)$, $x(\lambda p, \lambda w)$ and $x(p, w)$ are solutions to the same problem, hence they must be equal. □

Q.E.D.

The upshot of this result is that if prices go up by a factor λ, but so does wealth, the purchasing pattern of an economic agent will not change. Similarly, it does not matter whether prices and incomes are expressed in dollars, rupees, euros or yuan: demand is still the same.

PROPOSITION 5: (Walras' Law) *If preferences are locally non-satiated, then for any* (p, w) *and* $x \in x(p, w)$, $p \cdot x = w$.

Proof. By contradiction. Suppose $x \in x(p, w)$ with $p \cdot x < w$. Then there is some $\varepsilon > 0$ such that for all y with $||x - y|| < \varepsilon$, $p \cdot y < w$. But then by local non-satiation, there must be some bundle y for which $p \cdot y < w$ and $y \succ x$. Hence $x \notin x(p, w)$ — a contradiction. □

Q.E.D.

Walras's Law says that a consumer with locally non-satiated preferences will consume her entire budget. In particular, this allows us to re-express the consumer problem as:

Consumer Problem

$$\max_{x \in \mathbb{R}^n_+} u(x)$$
$$\text{such that } p \cdot x = w$$

where the budget inequality is replaced with an equality.

The next result speaks to our earlier observation that there might be many solutions to the consumer problem.

PROPOSITION 6: (Convexity/Uniqueness) *If preferences are convex, then $x(p, w)$ is convex-valued. If preferences are strictly convex, then the consumer optimum is always unique, that is, $x(p, w)$ is a singleton.*

Proof. Suppose preferences are convex and $x, x' \in x(p, w)$. For any $t \in [0, 1]$, $tx+(1-t)x' \in B(p, w)$ because $p\cdot(tx+(1-t)x') = tp\cdot x+(1-t)p\cdot x' \leq tw + (1-t)w = w$. Then, since $x \succeq x'$ and preferences are convex, we also have $tx + (1-t)x' \succsim x'$. Hence, $tx + (1-t)x' \in x(p, w)$. □

If preferences are strictly convex, the same construction leads to a contradiction. Suppose $x, x' \in x(p, w)$ with $x \neq x'$. Then strict convexity means that for any $t \in (0, 1)$, $tx + (1-t)x' \succ x'$. Hence, $x' \notin x(p, w)$. □

Q.E.D.

Thus, assuming the consumer's utility is continuous and locally non-satiated, we have established four properties of the Marshallian demand function $x(p, w)$:

1. it "exists",
2. is insensitive to proportional increases in price and income,
3. exhausts the consumer's budget, and
4. is single-valued if preferences are strictly convex.

The next result uses these properties to derive restrictions on the derivatives of the demand function.

PROPOSITION 7 (Adding Up) *Suppose preferences are locally non-satiated, and Marshallian demand is a differentiable function of prices and wealth. Then*

1. *A proportional change in all prices and income doesn't affect demand:* $\forall$ p, w, *and* $i = 1, \ldots, n$,

$$\sum_{j=1}^{n} p_j \frac{\partial}{\partial p_j} (x_i(p, w)) + w \frac{\partial}{\partial w} (x_i(p, w)) = 0$$

2. *A change in the price of one good won't affect total expenditure:* $\forall$ p, w, *and* $i = 1, \ldots, n$,

$$\sum_{j=1}^{n} p_j \frac{\partial}{\partial p_i} (x_j(p, w)) + x_i(p, w) = 0$$

3. *A change in income will lead to an identical change in total expenditure:* $\forall$ p, w,

$$\sum_{i=1}^{n} p_i \frac{\partial}{\partial w} (x_i(p, w)) = 1$$

Proof. (1) This follows directly from homogeneity. For all i, $x_i(\lambda p, \lambda w) = x_i(p, w)$ by homogeneity. Now differentiate both sides by λ and evaluate at $\lambda = 1$ to obtain the result. □

(2) This follows from Walras's Law. For non-satiated preferences, $p \cdot x(p, w) = w$ holds for all p and w. Differentiating both sides by p_i gives the result. □

(3) This also follows from Walras' Law. For non-satiated preferences, $p \cdot x(p, w) = w$ for all p, w. Differentiating both sides by w gives the result. □

Q.E.D.

Chapter 6

Indirect Utility

The *indirect utility function* $v(p, w)$ is defined as: Indirect Utility Function

$$v(p, w) = \max u(x) \text{ subject to } p \cdot x \leq w.$$

So $v(p, w)$ is the *value* of the consumer problem, or the most utility an agent can get at prices p with wealth w.

PROPOSITION 8: (Properties of v) *Suppose u is a continuous utility function representing a locally non-satiated preference relation* $\succsim$ *on* R^n_+*. Then* $v(p, w)$ *is:*

1. *homogeneous of degree zero:* $\forall\ p,\ w,$ *and* $\lambda > 0,\ v(\lambda p, \lambda w) = v(p, w)$
2. *continuous on* $\{(p, w) \mid p > 0,\ w \geq 0\}$
3. *non-increasing in* p *and strictly increasing in* w
4. *quasi-convex (i.e. the set* $\{(p, w) \mid v(p, w) \leq \bar{v}\}$ *is convex for any* $\bar{v}$*.*

Proof. (1) Homogeneity follows by a familiar argument. If we multiply both prices and wealth by a factor λ, the consumer problem is unchanged. $\square$

(2) Let $p^n \to p$ and $w^n \to w$ be sequences of prices and wealth. We must show that $\lim_{n\to\infty} v(p^n, w^n) = v(p, w)$, which we do by showing that $\liminf_n v(p^n, w^n) \geq v(p, w) \geq \limsup_n v(p^n, w^n) \geq \liminf_n v(p^n, w^n)$. The last inequality is true by definition, so we focus attention on the first two inequalities.

For the first inequality, let $x \in x(p, w)$, so that $v(p, w) = u(x)$ and let $a^n = w^n/(p^n \cdot x)$. Then, $a^n x \in B(p^n, w^n)$, so $v(p^n, w^n) \geq u(a^n x)$. By local non-satiation, $p \cdot x = w$, so $\lim_n a^n = \lim w^n/(p^n \cdot x) = w/(p \cdot x) = 1$. Hence, using the continuity of u, $\liminf_n v(p^n, w^n) \geq \liminf u(a^n x) = u(x) = v(p, w)$, which implies the first inequality.

For the second inequality, let $x^n \in x(p^n, w^n)$ so that $v(p^n, w^n) = u(x^n)$. Let n^k be a subsequence along which $\lim_{k\to\infty} u(x^{n^k}) = \limsup_n v(p^n, w^n)$. Since $p^n \gg 0$, the union of the budget sets defined by (p^{n^k}, w^{n^k}) and (p, w) is bounded above by the vector b whose ith component is $b_i = (\sup w^n)/(\inf p_i^n)$. Since the sequence $\{x^{n^k}\}$ is bounded, it has some accumulation point x. Since $p^{n^k} \cdot x^{n^k} \leq w^{n^k}$, it follows by taking limits that $p \cdot x \leq w$. Thus, $v(p, w) \geq u(x) = \lim_{k\to\infty} u(x^{n^k}) = \limsup_n v(p^n, w^n)$, which implies the second inequality. $\square$

(3) For the first part, note that if $p' > p$, then $B(p',\, w) \subset B(p,\, w)$, so clearly $v(p',\, w) \leq v(p,\, w)$. For the second part, suppose $w' > w$, and let $x \in x(p,\, w)$. By Walras' Law, $p \cdot x = w < w'$, so by a second application of Walras' Law, $x \notin x(p,\, w')$. Hence, there is some $x' \in B(p,\, w')$ such that $u(x') > u(x)$. $\square$

(4) Suppose that $v(p, w) \leq \overline{v}$ and $v(p', w') \leq \overline{v}$. For any $t \in [0, 1]$, consider (p^t, w^t) where $p^t = tp + (1-t)p'$ and $w^t = tw + (1-t)w'$. Let x be such that $p^t \cdot x \leq w^t$. Then, $w^t \geq p^t \cdot x = tp \cdot x + (1-t)p' \cdot x$, so *either* $p \cdot x \leq w$ *or* $p' \cdot x \leq w'$ or both. Thus, either $u(x) \leq v(p, w) \leq \overline{v}$ or $u(x) \leq v(p', w') \leq \overline{v}$, so $u(x) \leq \overline{v}$. Consequently, $v(p^t, w^t) = \max_{x: p^t \cdot x \leq w^t} u(x) \leq \overline{v}$. $\square$ *Q.E.D.*

Chapter 7

Demand with Derivatives

7.1 Lagrangian

How does one actually solve for Marshallian demand, given preferences, prices and wealth? If the utility function is differentiable,[1] then explicit formulae can sometimes be derived by analyzing the Lagrangian for the consumer problem:

$$\mathcal{L}(x, \lambda, \mu; p, w) = u(x) + \lambda [w - p \cdot x] + \sum_{i=1}^{n} \mu_i x_i$$

where λ is the Lagrange multiplier on the budget constraint and, for each i, μ_i is the multiplier on the constraint that $x_i \geq 0$. The "Lagrangian problem" is:

$$\min_{\lambda \geq 0, \mu \geq 0} \max_x \mathcal{L}(x, \lambda, \mu) = \min_{\lambda \geq 0, \mu \geq 0} \max_x u(x) + \lambda [w - p \cdot x] + \sum_{i=1}^{n} \mu_i x_i$$

[1] There are mixed opinions about the differentiability assumption. On one hand, there is no natural restriction on the underlying preference relation $\succeq$ that guarantees differentiability. Purists claim that makes the assumption of dubious validity. On the other, there exists no finite set of observed choices $C(X, \succsim)$ from finite sets X that ever contradicts differentiability. Pragmatists conclude from this that differentiability of demand is empirically harmless and can be freely adopted whenever it is analytically useful.

Kuhn-Tucker Conditions

The first order conditions for the maximization problem are:

$$\frac{\partial u}{\partial x_i} = \lambda p_i - \mu_i \,, \tag{7.1}$$

and $\lambda \geq 0$ and $\mu_i \geq 0$ for all $i = 1, \ldots, n$. The solution must also satisfy the original constraints:

$$p \cdot x \leq w \qquad \text{and} \qquad x \geq 0,$$

as well as the complementary slackness conditions:

$$\lambda\,(w - p \cdot x) = 0 \qquad \text{and} \qquad \mu_i x_i = 0 \text{ for } k = 1, ..., n.$$

These conditions, taken together are called the *Kuhn-Tucker conditions*.

Suppose we find a triplet (x, λ, μ) that satisfies the Kuhn-Tucker conditions. Does x solve the maximization problem, so that $x \in x(p, w)$? Conversely, if $x \in x(p, w)$ is a solution to the consumer problem, will x also satisfy the Kuhn-Tucker conditions along with some (λ, μ)?

The *Kuhn-Tucker Theorem* tells us that *if* $x \in x(p, w)$, then (subject to a certain "regularity" condition) there exist (λ, μ) such that (x, λ, μ) solve the Kuhn-Tucker conditions. Of course, there may be other solutions to the Kuhn-Tucker conditions that do not solve the consumer problem. However, if u is also quasi-concave and has an additional property, then the solutions to the consumer problem and the Kuhn-Tucker conditions coincide exactly, that is, the Kuhn-Tucker conditions are *necessary and sufficient* for x to solve the consumer problem.

Kuhn-Tucker Theorem

PROPOSITION 9 (Kuhn-Tucker Theorem) *Suppose that u is continuously differentiable and that $x \in x(p, w)$ is a solution to the consumer problem. If the constraint qualification holds at x, then there exists λ, $\mu_1, \ldots, \mu_n \geq 0$ such that (x, λ, μ) solve* *the Kuhn-Tucker conditions. Moreover, if u is quasi-concave and has the property that $[u(x') > u(x)] \Rightarrow [\nabla u(x) \cdot (x' - x) > 0]$, then any x that is part of a solution to the Kuhn-Tucker conditions is also a solution to the consumer problem.*

Proof. See the math review handout (starting on page 18).

Q.E.D.

7.2 Marginal Rate of Substitution

We can use the Kuhn-Tucker conditions to characterize Marshallian demand. First, using (7.1), we may write:

$$\frac{\partial u}{\partial x_i} \leq \lambda p_i, \quad \text{and if } x_i > 0 \ \frac{\partial u}{\partial x_i} = \lambda p_i.$$

From this, we derive the following important relationship: for all goods i and j consumed in positive quantity:

$$MRS_{ij} = \frac{\partial u(x(p,w))/\partial x_i}{\partial u(x(p,w))/\partial x_j} = \frac{p_i}{p_j}.$$

This says that at the consumer's maximum, *the marginal rate of substitution between i and j equals the ratio of their prices.* Marginal rate of Substitution

Were this not the case, the consumer could do better by marginally changing her consumption. For example, consider $x^* \in x(p,w)$. If $\frac{\partial u(x^*)/\partial x_1}{\partial u(x^*)/\partial x_2} > \frac{p_1}{p_2}$, then an increase in the consumption of good 1 of dx_1, combined with a decrease in the consumption of good 2 of $\frac{p_1}{p_2}\,dx_1$ would be feasible and would yield a utility change of $\frac{\partial u(x^*)}{\partial x_1}\,dx_1 - \frac{\partial u(x^*)}{\partial x_2}\,\frac{p_1}{p_2}\,dx_1 > 0$. Therefore, $x^* \notin x(p,w)$

Figures 7.1 and 7.2 give a graphical representation of the solution to the consumer problem. In Figure 7.1 both goods are consumed in positive quantities, so $\mu_1 = \mu_2 = 0$, and the marginal rate of substitution along the indifference curve equals the slope of the budget line at the optimum.

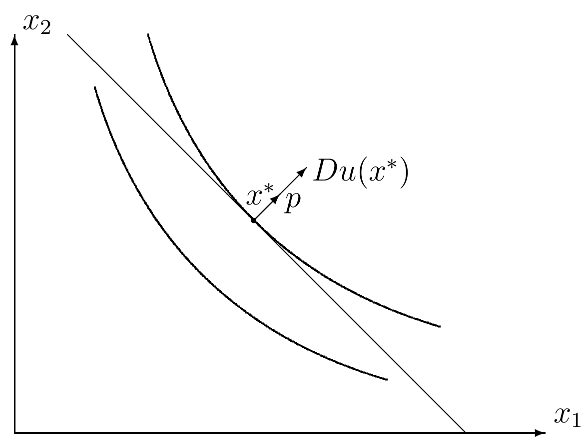

Figure 7.1: Marshallian demand: interior solution.

In Figure 7.2, we have a corner solution, so $\mu_1 > 0$ while $\mu_2 = 0$. Here the MRS does not equal the price ratio.

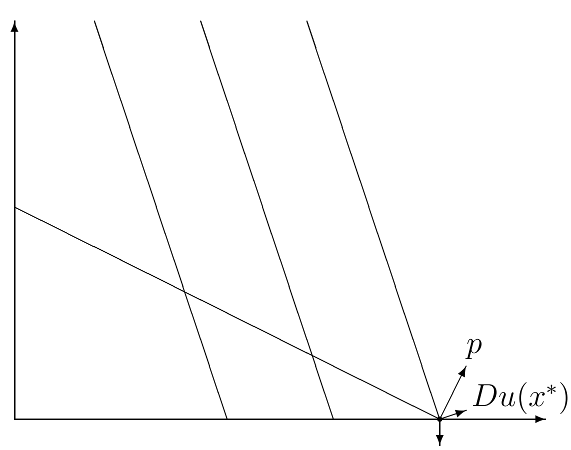

Figure 7.2: Marshallian demand: corner solution.

7.3 Lagrange Multiplier

The Lagrange multiplier λ in the first order conditions (equations 7.1) gives the marginal, or shadow, value of relaxing the constraint in the CP. Therefore, λ equals the consumer's marginal utility of wealth at the optimum. To see this directly, consider the case where $x(p, w)$ is a differentiable function and that $\{x_i(p, w)\}_{i=1}^{n} > 0 \; \forall i$. By the chain rule, the change in utility from a marginal increase in w is given by

$$\nabla u(x(p, w)) \, \frac{\partial x(p, w)}{\partial w} \,,$$

where $\nabla u(x) = [\partial u(x)/\partial x_1, \ldots, \partial u(x)/\partial x_n]$ and $\frac{\partial x(p,w)}{\partial w} = [\partial x_1(p, w)/\partial w, \ldots, \partial x_n(p, w)/\partial w]'$.

Substituting λp for $\nabla u(x(p, w))$, we have that the change in utility from a marginal increase in w is given by $\lambda p \, \frac{\partial x(p,w)}{\partial w}$. Now, notice that by Proposition 7 (the adding up theorem) number 3, $\sum_{i=1}^{n} p_i \, \frac{\partial}{\partial w} \left(x_i \left(p, \, w\right)\right) = 1$, so

$$\lambda p \, \frac{\partial x(p, w)}{\partial w} = \lambda \sum_{i-1}^{n} p_i \frac{\partial x_i(p, w)}{\partial w} = \lambda \,.$$

In general, if $u(.)$ is quasi-concave and there is a unique solution to the consumer problem then v is differentiable at $(p, \, w)$ and $\frac{\partial v(p, \, w)}{\partial w} = \lambda \geq 0$.

The Lagrange multiplier λ gives the value (in terms of utility) of having an additional unit of wealth. Because of this, λ is the sometimes called the *shadow price of wealth* or the *marginal utility of wealth (or income)*. In terms of the history of thought, the terms *marginal utility of income* or and *marginal utility of wealth* were important, because utilitarians thought that such considerations would guide the choice of public policies that redistribute of wealth or income. However, nothing in the consumer theory developed so far suggests any basis for using the marginal utility of income or wealth, as we have defined it, to guide redistribution policies.

Lagrange Multiplier

7.4 Roy's Identity

For our next calculations, it will be useful to have $\lambda > 0$. One might think that adding an assumption of local non-satiation would imply that strict inequality, since it certainly implies that v is increasing in w. However, neither local non-satiation nor any other condition on consumer preferences $\succsim$ is sufficient for the desired conclusion.[2] However, if there is everywhere at least one good j for which $\partial u/\partial x_j > 0$, then one can infer that $\partial v/\partial w \geq (\partial u/\partial x_j)/p_j > 0$.

PROPOSITION 10 Roy's Identity *Suppose that v is differentiable at (p, w), with $p_i > 0$, $w > 0$, and $\frac{\partial v}{\partial w} > 0$. Then $x(p, w)$ is a singleton and*

$$x_i(p, w) = -\frac{\partial v(p, w)/\partial p_i}{\partial v(p, w)/\partial w} \ .$$

Proof. The indirect utility function is given by $v(p, w) = u(x(p, w))$. If we differentiate this with respect to p_i we obtain, $\partial v(p, w)/\partial p_i = \sum_{j=1}^n \frac{\partial u(x)}{\partial x_j} \frac{\partial x_j}{\partial p_i}$. Substituting in the first order conditions for $\frac{\partial u(x)}{\partial x_j}$ we get $\partial v(p, w)/\partial p_i = \lambda \sum_{j=1}^n p_j \frac{\partial x_j}{\partial p_i}$. Now note that by Proposition 7 (the adding up theorem) number 2, $\sum_{j=1}^n p_j \frac{\partial}{\partial p_i} (x_j (p, w)) + x_i (p, w) = 0$. So the expression now becomes $\partial v(p, w)/\partial p_i = -\lambda x_i (p, w)$. As just shown in our discussion of Lagrange multipliers, $\frac{\partial v(p, w)}{\partial w} = \lambda$. Together, this establishes the identity. $\square$

[2] This is a purely technical point, but it serves to remind us that the same preferences can be represented in quite different ways. Suppose a representation u is selected so that the corresponding indirect utility satisfies $\frac{\partial v(p,w)}{\partial w} > 0$. Suppose $v(p, w) = \overline{v}$ and consider the alternative representation $\widehat{u}(x) = (u(x) - \overline{v})^3$. This utility function represents the same preferences as u and has corresponding indirect utility function $\widehat{v}(p, w) = \widehat{u}(x(p, w)) = (v(p, w) - \overline{v})^3$. Applying the chain rule leads to $\frac{\partial \widehat{v}(p,w)}{\partial w} = 0$. Thus, whether the marginal utility of income is positive or zero at a point is not just a property of the preferences themselves, but is a joint property of the preferences and their representation.

Q.E.D.

7.5 Homotheticity

DEFINITION: Homotheticity A function $f(x)$ is homothetic if $f(x) = g(h(x))$ where g is a strictly increasing function and h is a function that is homogeneous of degree 1, that is, a function is called homothetic if it is a positive monotonic transformation of a function that is homogeneous of degree 1.

Homotheticity

Economists often find it is useful to assume that utility functions are homothetic. Notice that this assumption is not much different than that of homogeneous of degree 1. Utility functions are only defined up to a monotonic transformation. Therefore, assuming that preferences can be represented by a homothetic function is equivalent to assuming that they can be represented by a function that is homogeneous of degree 1.

The importance of homothetic utility functions is that one indifference curve is much like another. Slopes of the curves depend only on the ratio of the goods and not on how far the curve is from the origin. We can study the behavior of an individual who has homothetic preferences by looking only at one indifference curve or at a few nearby curves without fearing that our results would change dramatically at very different levels of utility.

This can also be a useful way to identify a homothetic utility function. The marginal rate of substitution for a homothetic utility function depends only on the ratio of the amounts of the two goods and not on the individual quantities of the goods.

Chapter 8

Hicksian Demand

8.1 Expenditure Minimization Problem

We now make what will prove to be a very useful detour in consumer theory and introduce the consumer's *expenditure minimization problem* (EMP).

Expenditure Minimization Problem

$$\min_{x\in\mathbb{R}^n_+} p\cdot x$$
$$\text{s.t. } u(x)\geq u,$$

where $u\geq u(0)$ and $p_i>0\,\forall\, i$. This problem finds the cheapest bundle at prices p that yields utility of at least u.[1] The solution $h(p,u)$ of the expenditure minimization problem is called the *Hicksian* (or *compensated*) *demand*. The question now is, when does the EMP have a solution?

[1] This problem is sometimes called the "*dual consumer problem*," but that terminology suggests incorrectly that "duality" results always apply. In general, duality results will apply to this problem only when u is quasi-concave, but that property plays no role in most of our analysis. It is more accurate to refer to this problem as the "*expenditure minimization problem*."

Proposition 11 (Existence) *If $p_i > 0 \,\forall\, i$, $u(.)$ is continuous, and there is some x such that $u(x) \geq u$, then (*EMP*) has a solution.*

Proof. Let $u(\widehat{x}) \geq u$. Let $S = \{x | p \cdot x \leq p \cdot \widehat{x}\} \cap \{x | u(x) \geq u\}$. By construction, the first set in the intersection is non-empty and compact and by continuity of u, the second set is closed, so S is a compact set. Hence, the continuous function $p \cdot x$ achieves its minimum at some point x^* on S. By construction, for every $x \notin S$ such that $u(x) \geq u$, $p \cdot x > p \cdot \widehat{x} \geq p \cdot x^*$, so x^* solves the EMP.□

Q.E.D.

Expendlture Function

We define the *expenditure function* to be the corresponding value function:

$$e(p, u) = \min_{x \in \mathbb{R}^n_+} p \cdot x \text{ subject to } u(x) \geq u\,.$$

Thus, $e(p, u)$ is the minimum expenditure required to achieve utility u at prices p, and $h(p, u)$ is the set of consumption bundles that the consumer would purchase at prices p if she wished to minimize her expenses but still achieve utility u.

What is the motivation for introducing the expenditure minimization problem, when we have already analyzed the "actual" consumer problem? We take this detour to capture two main advantages.

1. We will use the expenditure function to decompose the effect of a price change on Marshallian demand into two corresponding effects. On one hand, a price reduction makes the consumer wealthier, just as if she had received a small inheritance, and that could certainly affect demand for all goods. We will call that the *wealth effect* (or *income effect*). In addition, even if the consumer were forced to give up her extra wealth,

the price reduction would cause the optimizing consumer to substitute the newly cheaper good for more expensive ones and perhaps to make other changes as well. That is called the *substitution effect.*

2. The expenditure function also turns out to play a central role in welfare economics. More about that later in these notes.

With these advantages lying ahead, we first introduce three propositions to identify, respectively, the properties of Hicksian demand, those of the expenditure function, and the relationship between the two functions.

Hicksian Demand Properties

PROPOSITION 12 (Properties of Hicksian Demand) *Suppose $u(.)$ is a continuous utility function representing a preference relation $\succsim$ on $\mathbb{R}^n_+$. Then*

1. *Homogeneity: $h(p,u)$ is homogeneous of degree zero in p. For any p, u, and $\lambda > 0$, $h(\lambda p, u) = h(p,u)$.*
2. *No Excess Utility: If $u \geq u(0)$ and $p_i > 0 \,\forall\, i$, then $\forall\, x \in h(p,u)$, $u(x) = u$.*
3. *Convexity/Uniqueness: If preferences are convex, then $h(p,u)$ is a convex set. If preferences are strictly convex and $p_i > 0 \,\forall\, i$, then $h(p,u)$ is a singleton.*

Proof. (1) Note that the constraint set, or choice set, is the same in the expenditure problem for $(\lambda p, u)$ and (p,u). But then

$$\min_{\left\{x \in \mathbb{R}^n_+ : u(x) \geq u\right\}} \lambda p \cdot x = \lambda \min_{\left\{x \in \mathbb{R}^n_+ : u(x) \geq u\right\}} p \cdot x,$$

so the expenditure problem has the same solution for $(\lambda p, u)$ and (p,u). $\square$

(2) Suppose to the contrary that there is some $x \in h(p,u)$ such that $u(x) > u \geq u(0)$. Consider a bundle $x' = tx$ with $0 < t < 1$. Then $p \cdot x' < p \cdot x$, and by the intermediate value theorem, there is some t such that $u(x') \geq u$, which contradicts the assumption that $x \in h(p,u)$. □

(3) Note that $h(p,u) = \{x \in \mathbb{R}^n_+ \mid u(x) \geq u\} \cap \{x \mid p \cdot x = e(p,u)\}$ is the intersection of two convex sets and hence is convex. If preferences are strictly convex and $x, x' \in h(p,u)$, then for $t \in (0,1)$, $x'' = tx + (1-t)x'$ satisfies $x'' \succ x$ and $p \cdot x'' = e(p,u)$, which contradicts "no excess utility." □

Q.E.D.

Expenditure Function Properties

Proposition 13 Properties of the Expenditure Function *Suppose u is a continuous utility function representing a locally non-satiated preference relation $\succsim$ on $\mathbb{R}^n_+$. Then $e(p,u)$ is*

1. *Homogeneous of degree one in p: $\forall\, p$, u and $\lambda > 0$, $e(\lambda p, u) = \lambda e(p,u)$*
2. *Continuous in p and u*
3. *Nondecreasing in p and strictly increasing in u provided $p_i > 0 \,\forall\, i$*
4. *Concave in p.*

Proof. (1) As in proposition 12, note that

$$e(\lambda p, u) = \min_{\{x \in \mathbb{R}^n_+ : u(x) \geq u\}} \lambda p \cdot x \;=\; \lambda \min_{\{x \in \mathbb{R}^n_+ : u(x) \geq u\}} p \cdot x \;=\; \lambda e(p,u).$$

So, $e(p,u)$ is homogeneous of degree 1. □

(2) I omit this proof, which is similar to proving continuity of the indirect utility function. □

(3) Let $p' > p$ and suppose $x \in h(p',u)$. Then $u(h) \geq u$, and $e(p',u) = p' \cdot x \geq p \cdot x$. It follows immediately that $e(p,u) \leq e(p',u)$. Therefore, $e(p,u)$ is nondecreasing in p. □

For the second statement, suppose to the contrary that $e(p,u)$ is nondecreasing in u, that is $u' > u$ with $e(p,u') \leq e(p,u)$. Then, for some $x \in h(p,u)$, $u(x) = u' > u$, which contradicts the "no excess utility" conclusion of proposition 12. □

(4) Fix u, let $p'' = tp + (1-t)p'$ for $t \in (0,1)$, and suppose $x \in h(p'',u)$. Then, $p \cdot x \geq e(p,u)$ and $p' \cdot x \geq e(p',u)$, so $e(p'',u) = (tp + (1-t)p') \cdot x = t(p \cdot x) + (1-t)(p \cdot x) \geq te(p,u) + (1-t)e(p',u)$. Therefore, $e(p,u)$ is concave in p. □

Q.E.D.

A natural question here is how are the Hicksian demands related to the expenditure function. For an answer to this question, we turn to Shepard's Lemma.

Shepard's Lemma

Proposition 14: Shepard's Lemma *Suppose that $u(.)$ is a continuous utility function representing a locally non-satiated preference relation $\succsim$ and suppose that $h(p,u)$ is a singleton. Then the expenditure function is differentiable in p, and for all $i = 1, \ldots, n$,*

$$\frac{\partial e(p,u)}{\partial p_i} = h_i(p,u)\,.$$

Proof. Using the chain rule, the change in expenditure can be written as:

$$\frac{\partial e(p,u)}{\partial p_i} = \frac{\partial (p_i h(p,u))}{\partial p_i} = h(p,u) + p_i \frac{\partial h(p,u)}{\partial p_i}\,.$$

From the first order conditions for an interior solutions to the EMP, $p_i = \lambda \frac{\partial u(h(p,u))}{\partial x_i}$. So substituting into the expression above:

$$\frac{\partial e(p,u)}{\partial p_i} = h(p,u) + \lambda \frac{\partial u(h(p,u))}{\partial x_i} \frac{\partial h(p,u)}{\partial p_i}\,.$$

But, since the constraint $u(h(p,u)) = u$ holds for all p in the EMP, we know that $\frac{\partial u(h(p,u))}{\partial x_i}\frac{\partial h(p,u)}{\partial p_i} = 0$. Therefore, $\frac{\partial e(p,u)}{\partial p_i} = h(p,u)$. □

Q.E.D.

Thought question: What is $\partial e/\partial u$?

8.2 Hicksian Comparative Statics

Comparative statics are statements about how the solution to a problem will change with parameters. In the consumer problem, the parameters are (p, w), so comparative statics are statements about how $x(p, w)$, or $v(p, w)$ will change with p and w. Similarly, in the expenditure problem, the parameters are (p, u), so comparative statics are statements about how $h(p, u)$ or $e(p, u)$ will change with p and u.

Our first result gives a comparative statics statement about how a change in price changes the expenditure required to achieve a given utility level u. The "law of demand" formalizes to the idea when the price of some good increases, the (Hicksian) demand for that good decreases.

Proposition 15: Law of Hicksian Demand *Suppose* p, p' ≥ 0 *and let* $x \in h(p,u)$ *and* $x' \in h(p',u)$. *Then,* $(p'-p)(x'-x) \leq 0$.

Proof. By definition $u(x) \geq u$ and $u(x') \geq u$. So, by optimization, $p' \cdot x' \leq p' \cdot x$ and $p \cdot x \leq p \cdot x'$. We may rewrite these two inequalities as $p' \cdot (x' - x) \leq 0$ and $0 \geq -p \cdot (x' - x)$, and the result follows immediately. □

Q.E.D.

The Law of Demand can be applied to study how demand for a single good varies with its own price. Thus, suppose that the only difference between p'

and p is that, for some k, $p'_k > p_k$, but $p'_i = p_i$ for all $i \neq k$. Then, with single-valued demand,

$$(p'_k - p_k)\,[h_k(p', u) - h_k(p, u)] \leq 0\,.$$

This means that $h_k(p, u)$ is *decreasing* in p_k. Or in words, *Hicksian demand curves slope downward.*

A simple way to see this graphically is to note that the change in Hicksian demand given a change in price is a shift along an indifference curve:

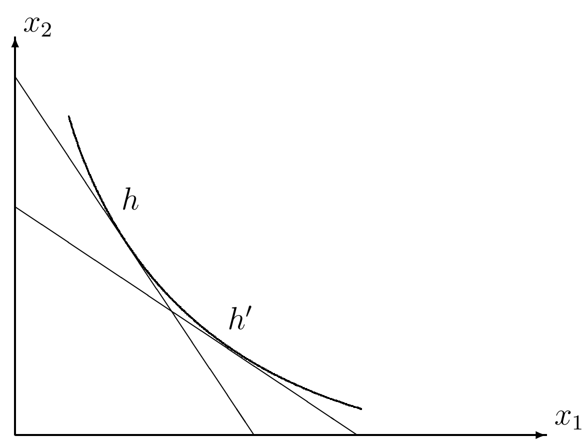

Figure 8.1: Hicksian Demand: Change in Price

In contrast, Marshallian demand $x_k(p, w)$ need not be decreasing in p_k (though this is *typically* the case). To see why, consider Figure 8.2. We will come back to how Marshallian demand reacts to a change in price, and to the relationship between the change in Marshallian and Hicksian demand, in a minute.

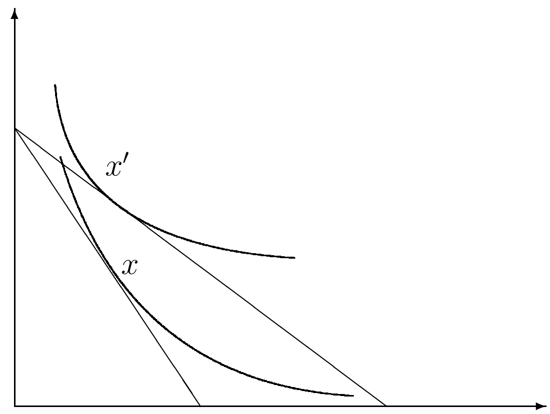

Figure 8.2: Marshallian demand may increase with price increase.

If the Hicksian demand function $h(p, u)$ is singleton-valued and continuously differentiable, we can use derivatives to describe how this demand responds to price changes. The next result is closely related to corresponding results that we have previously discussed concerning a firm's input demands.

Consider the matrix:

$$D_p h(p,u) = \begin{pmatrix} \frac{\partial h_1(p,u)}{\partial p_1} & \frac{\partial h_n(p,u)}{\partial p_1} \\ \cdots & \cdots \\ \frac{\partial h_1(p,u)}{\partial p_n} & \frac{\partial h_n(p,u)}{\partial p_n} \end{pmatrix}.$$

Recall the definition that an $n \times n$ symmetric matrix D is *negative semi-definite* if for all $z \in \mathbb{R}^n$, $z \cdot Dz \leq 0$.

Proposition 16 *Suppose that u(.) represents a preference relation $\succsim$ and that $h(p, u)$ a singleton and is continuously differentiable at (p, u), where $p_i > 0 \, \forall \, i$. Then $D_p h(p, u) = \frac{\partial h(p,u)}{\partial p}$ is symmetric and negative semi-definite.*

Proof. By Shephard's Lemma, $h_i(p, u) = \frac{\partial e(p,u)}{\partial p_i}$, so $\frac{\partial h_i(p,u)}{\partial p_j} = \frac{\partial^2 e(p,u)}{\partial p_i \partial p_j}$ We may rewrite this as:

$$D_p h(p, u) = D_p^2 e(p, u).$$

For symmetry, recall that Young's Theorem from calculus tells us that for any twice continuously differentiable function $f(x, y)$, $f_{xy} = f_{yx}$. Applying this result shows us that

$$\frac{\partial h_i(p, u)}{\partial p_j} = \frac{\partial^2 e(p, u)}{\partial p_i \partial p_j} = \frac{\partial^2 e(p, u)}{\partial p_j \partial p_i} = \frac{\partial h_j(p, u)}{\partial p_i}$$

and hence $D_p(p, u)$ is symmetric. □

For negative semi-definiteness, recall the definition that an $n \times n$ symmetric matrix D is *negative semi-definite* if for all $z \in \mathbb{R}^n$, $z \cdot Dz \leq 0$. From proposition 13, $e(p, u)$ is a concave function of p. This implies that $D_p^2 e(p, u)$ is negative semi-definite (see the appendix D of MWG for a proof.) □

Q.E.D.

What is most surprising here is the symmetry of the demand matrix: the effect of a small increase in the price of good i on the quantity demanded of good j is identical to effect of a similar increase in the price of good j on the quantity demand of good i. Thus, the derivative of the Hicksian demand for butter, say in kilograms, with respect to the price of compact disks, say in \$/disk, is the same as the derivative of the Hicksian demand for compact disks with respect to the price of butter in \$/kilogram.

The proposition also encompasses within it a differential form of the law of demand. For, the rate of change of the Hicksian demand for good j as the price p_j increases is $\partial h_j / \partial p_j = \partial^2 e(p, u) / \partial p_j^2$, which is a diagonal element of the matrix $D_p h$. The diagonal elements of a negative semi-definite matrix are always non-positive. To see why, let $z = (0, ...0, 1, 0, ..., 0)$ have a 1 only in its j^{th} place. Then since $D_p h(p, u)$ is negative semi-definite, $0 \geq z D_p h(p, u) \cdot z = \partial h_j / \partial p_j$, proving that the j^{th} diagonal element is non-positive. That is, the Hicksian demand for good j is weakly decreasing in the price of good j.

Chapter 9

The Slutsky Equation

9.1 Relating Hicksian & Marshallian Demand

Next, we bring the theory together by relating Marshallian and Hicksian demand and using that relationship to derive the Slutsky equation, which decomposes the effect of price changes on Marshallian demand.

Proposition 17 *Suppose u is a utility function representing a continuous, locally non-satiated preference relation $\succsim$ on R^n_+ and let $p_i > 0 \,\forall\, i$. Then,*

1. *For all $p > 0$ and $w \geq 0$, $h(p,\, v(p,w)) = x(p,w)$ and $e(p,\, v(p,w)) = w$*
2. *For all $p > 0$ and $u \geq u(0)$, $x(p,\, e(p,u)) = h(p,u)$ and $v(p,\, e(p,u)) = u$*

Proof. (1) Fix prices $p_i > 0 \,\forall\, i$ and wealth $w \geq 0$ and let $x \in x(p,w)$. Since $u(x) = v(p,w)$ and $p \cdot x \leq w$, it follows that $e(p, v(p,w)) \leq w$. For the reverse inequality, we use the hypothesis of local non-satiation. It implies Walras'

Law, so for any x' with $p \cdot x' < w$, it must be that $u(x') < v(p, w)$. So, $e(p, v(p, w)) \geq w$. Combining these implies that $e(p, v(p, w)) = w$ and hence that $h(p, v(p, w)) = x(p, w)$. □

(2) Fix prices $p_i > 0 \, \forall \, i$ and target utility $u \geq u(0)$ and let $x \in h(p, u)$. Since $u(x) \geq u$, it follows that $v(p, e(p, u)) \geq u(x) \geq u$. By the "no excess utility" proposition, for any x' with $u(x') > u$, $p \cdot x' > p \cdot x = e(p, u)$. Thus, $v(p, e(p, u)) \leq u$. So, $v(p, e(p, u)) = u$ and it follows that $x(p, e(p, u)) = h(p, u)$. □

Q.E.D.

This result is quite simple and intuitive (at least after one understands the local non-satiation condition). It says that if $v(p, w)$ is the most utility that a consumer can achieve with wealth w at prices p, then to achieve utility $v(p, w)$ will take wealth at least w. Similarly, if $e(p, u)$ is the amount of wealth required to achieve utility u, then the most utility a consumer can get with wealth $e(p, u)$ is exactly u.

9.2 The Slutsky Decomposition

Proposition 18 Slutsky Equation *Suppose u is a continuous utility function representing a locally non-satiated preference relation $\succsim$ on $\mathbb{R}^n_+$ and let $p_i > 0 \, \forall \, i$ and $w = e(p, u)$. If $h(p, u)$ and $x(p, w)$ are singleton-valued and differentiable at (p, u, w), then for all i, j,*

$$\frac{\partial x_i(p, w)}{\partial p_j} = \frac{\partial h_i(p, \bar{u})}{\partial p_j} - \frac{\partial x_i(p, w)}{\partial w} x_j(p, w).$$

Proof. Starting with the identity,

$$h_i(p, \overline{u}) = x_i(p, e(p, \overline{u}))$$

letting $w = e(p, u)$ and differentiating with respect to p_j gives:

$$\frac{\partial h_i(p, \overline{u})}{\partial p_j} = \frac{\partial x_i(p, w)}{\partial p_j} + \frac{\partial x_i(p, w)}{\partial w} \frac{\partial e(p, \overline{u})}{\partial p_j}.$$

Substituting in for the last term using Shephard's lemma$\left(\frac{\partial e(p,u)}{\partial p_i} = h_i(p, u)\right)$ and the identity $h_i(p, \overline{u}) = x_i(p, w)$ gives the result. □

Q.E.D.

The Slutsky equation is interesting for two reasons. First, it gives a (fairly simple) relationship between the Hicksian and Marshallian demands. More importantly, it allows us to analyze the response of Marshallian demand to price changes, breaking it down into two distinct effects:

$$\underbrace{\frac{\partial x_i(p, w)}{\partial p_i}}_{\text{total effect}} = \underbrace{\frac{\partial h_i(p, \overline{u})}{\partial p_i}}_{\text{substitution effect}} - \underbrace{\frac{\partial x_i(p, w)}{\partial w} x_i(p, w)}_{\text{wealth effect}}$$

An increase in p_i does two things. It causes the consumer to *substitute* away from i toward other relatively cheaper goods. And second, it makes the consumer poorer, and this *wealth effect* also changes his desired consumption — potentially in a way that counteracts the substitution effect.

The Slutsky equation contains within it a suggestion about how to test the subtlest prediction of consumer choice theory. Given enough data about $x(p, w)$, one can derive the matrix of derivatives $D_p x$ and add to each term the corresponding wealth effect to recover the matrix of substitution effects, which corresponds to $D_p h$. If consumers are maximizing, then the matrix obtained in that way must be symmetric (and negative semi-definite).

Many economists have regarded this analysis and its symmetry conclusion as a triumph for the use of formal methods in economics. The analysis does demonstrate the possibility of using theory to derive subtle, testable implications that had been invisible to researchers using traditional verbal and graphical methods. Historically, that argument was quite influential, but its influenced has lessened over time. Critics typically counter it by observing that formal research has generated few such conclusions and that the maximization hypothesis on which all are based fares poorly in certain laboratory experiments.

Figure 9.1 illustrates the Slutsky equation, decomposing the demand effect of a price change into substitution and wealth effects. Fixing wealth w, when the price drops from $p = (p_1, p_2)$ to $p' = (p_1', p_2)$ with $p_1' < p_1$, the demand changes from x to x'. Letting $u = v(p, w)$ and $u' = v(p', w)$, note that $x = h(p, u)$, and $x' = h(p, u')$. Then the shift from x to x' can be decomposed as follows. The *substitution effect* is the consumer's shift along her indifference curve from $x = h(p, u)$ to $h(p', u)$ and a *wealth effect* or *income effect* is the consumer's shift from $h(p', u)$ to $x(p', w)$. Why is this second effect a wealth effect? Because $h(p', u) = x(p', e(p', u))$ the move corresponds to the change in demand at prices p' from increasing wealth from $e(p', u)$ to $w = e(p', u')$.

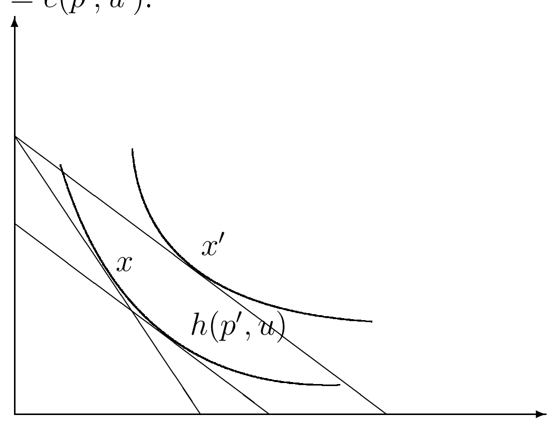

Figure 9.1: Wealth and Substitution Effects

9.3 Demand Relationships Among Goods

Consider the following definitions:

Definition: Normal/Inferior Good Good i is a normal good if $x_i(p, w)$ is increasing in w. Good i is an inferior good if $x_i(p, w)$ is decreasing in w.

Definition: Regular/Giffen Good Good i is a regular good if $x_i(p, w)$ is decreasing in p_i. Good i is a Giffen good if $x_i(p, w)$ is increasing in p_i.

Definition: Substitute/Complement Good i is a substitute for good j if $h_i(p, u)$ is increasing in p_j. Good i is a complement for good j if $h_i(p, u)$ is decreasing in p_j.

Definition: Gross Substitute/Complement Good i is a gross substitute for good j if $x_i(p, w)$ is increasing in p_j. Good i is a gross complement for good j if $x_i(p, w)$ is decreasing in p_j.

Figure 9.2 shows what happens to Marshallian demand when prices change. Here, as the price of the first good decreases, Marshallian demand shifts from x to x' to x''. In this picture, the first good is regular — as its price decreases, the demand for it increases. Note also that as the price of good one decreases, the Marshallian demand for the second good also increases: so goods i and j are gross complements.

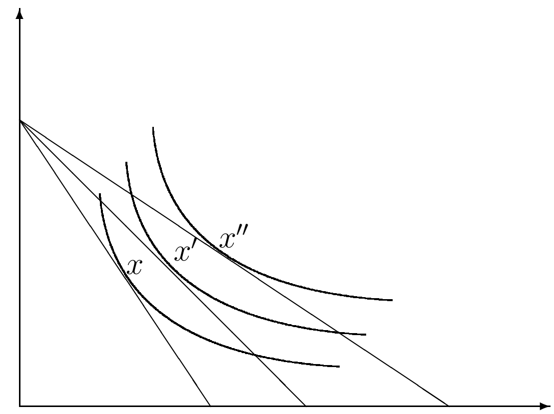

Figure 9.2: Offer Curves or Price Expansion Path

Traditional economics textbooks call a pair of goods substitutes or complements based on their Hicksian demands and reserve the terms "gross substitutes" and "gross complements" for the relations based on Marshallian demands. Perhaps the reason for this is that the Hicksian language is easier, because the Hicksian substitutes condition is a symmetric one, so one can simply say that "goods i and j are substitutes" without needing to specify which is a substitute for the other. The condition is symmetric because, as previously shown:

$$\frac{\partial h_i(p,u)}{\partial p_j} = \frac{\partial h_j(p,u)}{\partial p_i}.$$

In contrast, the gross substitute condition is not generally symmetric, because the wealth effect on $x_i(p,w)$ caused by an increase in p_j is not generally the same as the wealth effect on $x_j(p,w)$ caused by an increase in p_i:

$$\frac{\partial x_i(p,w)}{\partial w} x_j(p,w) \neq \frac{\partial x_j(p,w)}{\partial w} x_i(p,w).$$

In common practice, when one says that two or more goods are "gross substitutes," one means that each good is a gross substitute for each other good.

Keep in mind that even if goods are substitutes in one range of prices, they may still be complements for another range. When tight logical arguments are required, best practice is to describe assumptions in precise mathematical terms and to use terms like substitutes and complements as ways to describe and explicate the precise formal argument. Also, note that in a two-good world, the goods can not be complements, they must be substitutes. However, the two goods could be either gross complements or gross substitutes depending on the magnitude of the wealth effect.

Question: Suppose that good i is a complement for good j and that good i is a normal good. If p_j increases, will the consumer increase or decrease her consumption of good i?

Answer: Because good i is a normal good, the wealth effect is negative (an increase in the price of good j makes the consumer poorer). Because good i is a complement for good j, the substitution effect is negative. Both the wealth effect and the substitution effect are negative so we get an unambiguous sign for the total effect: $x_i(p, w)$ is decreasing in p_j so good i is a gross complement for good j. The consumer will decrease her consumption of good i for an increase in p_j.

Question: Suppose that good i is a substitute for good j and that good i is a normal good. If p_j increases, will the consumer increase of decrease her consumption of good i?

Answer: Because good i is a normal good, the wealth effect is negative. Because good i is a substitute for good j, the substitution effect is positive. This means that we get an ambiguous sign for the total effect, so we do not know if goods i or j are complements or substitutes. To answer this question we would need to know the magnitudes of the two effects.

9.4 Engel Curves

As a consumer's wealth increases, it is natural to expect that the quantity of each good purchased will also increase. This situation is illustrated in figure 9.3. As income increases from w to w' to w'', the budget line shifts out and Marshallian demand increases from x to x' to x''. Notice that the budget lines are all parallel, reflecting the fact that only wealth is changing and that prices have been held constant.

In figure 9.3, both goods increase as wealth increases, hence, both goods are normal. This is the usual situation and that is why we refer to these types of goods as "normal." However, for some goods the quantity chosen may decrease as wealth increases in some ranges. We call these goods "inferior." It is important to note that a good may be normal over a certain range and then inferior over another.

If we plot $x(p, w)$ for each possible income level w, and connect the points, the resulting curve is called an *Engel curve* or *Income expansion curve.*[1]

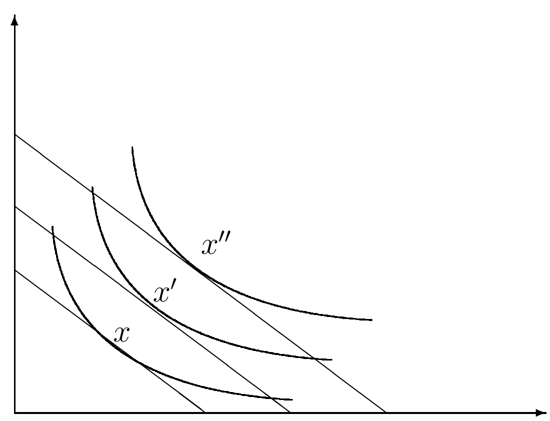

Figure 9.3: Engel Curve or Income Expansion Path

[1] An important question in development economics is how to estimate these curves empirically. The basic approach is to estimate the demands for major budget items — food, shelter, clothing — as a function of prices and income, and then ask how these demands have changed and will change as the country becomes richer.

Chapter 10

Consumer Welfare: Price Changes

10.1 Compensating and Equivalent Variation

We now turn to a particularly beautiful part of consumer theory: the measurement of consumer welfare. We assume throughout that consumer preferences are locally non-satiated and investigate the question: how much better or worse off is the consumer as the result of a change in prices from p to p' ?

Consumer Welfare

This question is much less narrow than it may seem. For purposes of determining welfare effects, many changes in the economic environment can be viewed as price changes. Taxes and subsidies are obvious cases: they add to or subtract from the price someone pays for a good. If we want to study the welfare effects of technical change, such as the introduction of a new product, we can formulate that as a change in price from $p = \infty$ to some finite price p'.

Let (p, w) be the consumer's status prior to the price change, and (p', w) the consumer's status after the price change. A natural candidate for measuring the change in welfare is to look at the change in the consumer's utility, i.e. at $v(p', w) - v(p, w)$. Of course, the problem is that this measure depends on *which* utility function we choose to represent the consumer preferences. While all give the same qualitative answer to the question of whether the consumer is better or worse off, they give different answers to the question of by how much she is better or worse off. In addition, the answer they give is in utils, which have no real meaning.

While there is no complete solution to this problem, there is an elegant partial solution. We can use the expenditure function to measure welfare changes in dollars. Essentially, we ask: *how much money is required to achieve a certain level of utility before and after the price change?* To answer this, we need to choose a level of utility as a reference point for making this comparison. There are two obvious candidates: the level of utility achieved by the consumer prior to the change and the level achieved after the change. We refer to these two measures as *compensating and equivalent variation.* Both are constructed to be positive for changes that increase welfare and negative for changes that reduce welfare.

Compensating variation specifies how much less wealth the consumer needs to achieve the same maximum utility at prices p' as she had before the price change. Letting $u = v(p, w)$ be the level of utility achieved prior to the price change,

Compensating Variation

$$\text{Compensating Variation } = e(p, u) - e(p', u) = w - e(p', u)\,.$$

That is, if prices change from p to p', the magnitude of compensating variation tells us how much we will have to charge or compensate our consumer to have her stay on the same indifference curve.

Equivalent variation gives the change in the expenditure that would be required at the original prices to have the same ("equivalent") effect on consumer as the price change had. Letting $u' = v(p', w)$ be the level of utility achieved after the price change.

Equivalent Variation

$$\text{Equivalent Variation } = e(p,u') - e(p',u') = e(p,u') - w.$$

That is, equivalent variation tells us how much more money the consumer would have needed yesterday to be as well off as she is today.

Figure 10.1 illustrates compensating variation for a situation where only a single price — that of the first good — changes. In this figure, think of the second good as a composite good (i.e. expenditures on all other items) measured in dollars[1]. Prices change from p to p' where $p_1' > p_1$ and $p_2' = p_2 = 1$ and the budget line rotates in. To identify compensating variation, we first find the wealth required to achieve utility u at prices p', i.e. $e(p',u)$, then find the difference between this and $w = e(p,u)$, the starting level of wealth.

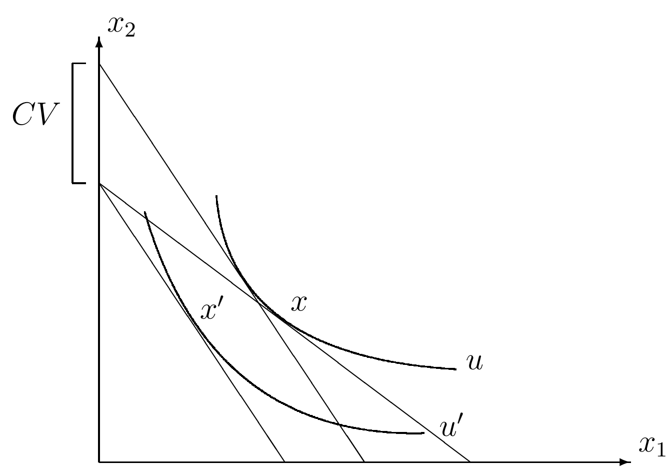

Figure 10.1: Compensating Variation

Figure 10.2 displays equivalent variation for the same hypothetical price

[1]Formally, this means we are working with the two-argument utility function $\hat{u}(x_1, y) = \max_{(x_2,\ldots,x_n)\in\mathbb{R}_+^{n-1}} u(x_1, \ldots x_n)$ subject to $p_2x_2 + \ldots + p_nx_n = y$.

change. Here, the first step is to find the wealth required to achieve utility u' at the original prices p.

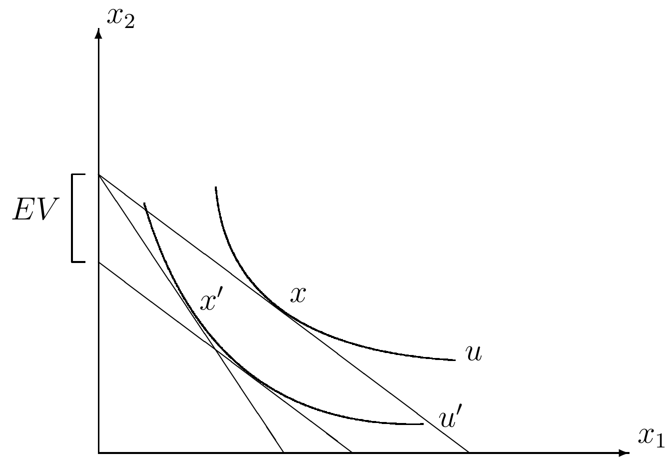

Figure 10.2: Equivalent Variation

Generally speaking, compensating and equivalent variation will not be the same, since they are the answers to different questions. There is, however, one case for which they will coincide. If preferences are quasi-linear, then equivalent and compensating variation are identical. Demonstrating this is left as a homework assignment.

If the price change affects only a single good i, we can relate equivalent and compensating variation to the Hicksian demand in this simple way:

$$\text{CV} = e(p,u) - e(p',u) = \int_{p_i'}^{p_i} \frac{\partial e(p,u)}{\partial p_i} dp_i = \int_{p_i'}^{p_i} h_i(p,u) dp_i \,,$$

and similarly

$$\text{EV} = e(p,u') - e(p',u') = \int_{p_i'}^{p_i} \frac{\partial e(p,u')}{\partial p_i} dp_i = \int_{p_i'}^{p_i} h_i(p,u') dp_i \,.$$

Figure 10.3 shows the Hicksian demand curves for a single good (good one) at two utility levels $u > u'$, assuming that the good is normal. To identify CV, we need to integrate the area to the left of the $h_1(\cdot, u)$ curve between p_1 and p_1'. Similarly, EV corresponds to the area to the left of the $h_1(\cdot, u')$ between p_1 and p_1'.

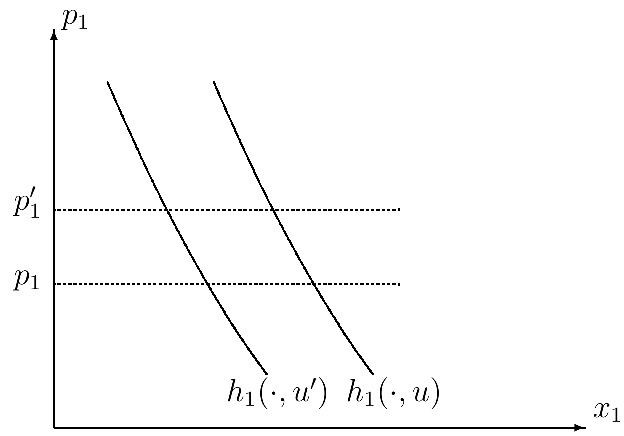

Figure 10.3: Relating Welfare to Demand

By construction, $w = e(p, u) = e(p', u')$. This allows us to conclude that $x(p, w) = h(p, u)$ and $x(p', w) = h(p', u')$. This relation is plotted in Figure 10.4.

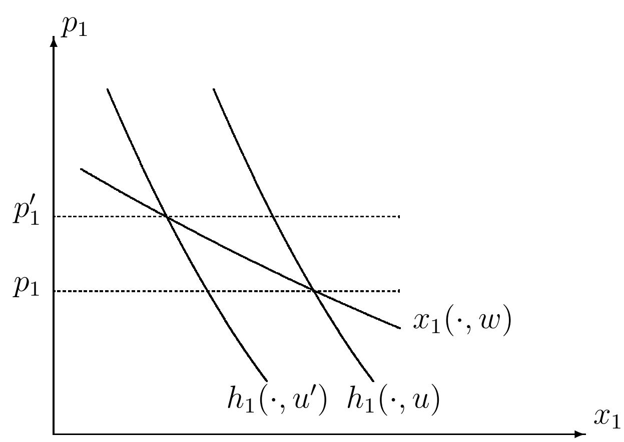

Figure 10.4: Relating Welfare to Demand

Figure 10.4 suggests that another measure of consumer welfare might be obtained by integrating to the left of the *Marshallian demand curve.* We define this — the area to the left of the Marshallian demand curve — as *consumer surplus.*

Consumer Surplus

$$\text{Consumer Surplus} = \int_{p_i'}^{p_i} x_i(p,w)dp_i\,.$$

In empirical work, where the regressions typically provide direct estimates of a Marshallian demand curve, Marshallian consumer surplus is a very common measure of consumer welfare. There is a long-standing debate in industrial organization as to when Marshallian Consumer Surplus is a good welfare measure (with important papers by Willig (1976, *AER)* and Hausman (1981, *AER)*). Consumer surplus has an important drawback — it does not have an immediate interpretation in terms of utility theory, as do EV and CV. However, one nice feature — which is apparent in the figure — is that Consumer Surplus is typically an intermediate measure that lies *between* compensating and equivalent variation. More precisely, on any range where the good in question is either normal or inferior,[2] we have the following relationship:

$$\min\{CV, EV\} \leq CS \leq \max\{CV, EV\}.$$

This typical relationship is sometimes used to justify consumer surplus as a welfare measure.

The most problematic part of using these concepts–equivalent variation, compensating variation, and consumer surplus–is the practice of simply adding up these numbers across individuals to compare overall welfare from two policies. Taken literally, this practice implies that one should be indifferent, in terms of overall welfare, between policies that redistribute benefits from

[2]These are not the only logical possibilities: it is also possible that the good is normal on part of the relevant domain and inferior on another part of the domain. Only in that case can the inequality fail.

the rich to the poor or from the poor to the rich. Because they omit distributional issues, these various measures can do no more than give an index of the equivalent or compensating changes in the total wealth of a society.

10.2 Consumer Welfare: Price Indices

In practice, perhaps the most important problem in the measurement of consumer welfare is obtaining correct measures of the growth of the economy. There are many subtleties involved in this measurement, depending on the goods that one includes in deciding about welfare. For example, there are important questions about how to measure public goods including environmental amenities, safety, and so on. One subtle issue concerns how to adjust for changes in the cost of living. That is, suppose one wants to know how much better off people are from one year to the next, given that economic growth has increased people's incomes (that is, has increased GDP). Once we have measured this increase in income, we need to account also for any changes in prices over the same period. Thus, to measure growth in consumer welfare, we need to "adjust" nominal income by a measure of the cost of living and use this adjusted measure (of "real income") to calculate growth.

This brings us to the topic of price indices. To define a price index (and this is essentially what the BLS does to measure inflation), one defines a "market basket" of goods — goods $1, 2, ..., n$ — and then compare their prices from period to period (quarterly, yearly, whatever). Let p be the prices of these goods "before"and p' the prices "after". There are basically two well–known ways to proceed. One way is to look at the quantities of the goods purchased in the "before" period, x, and compare the price of this basket at the two price levels. This is called a *Laspeyres index*:

$$\text{Laspeyres Index } = \frac{p' \cdot x}{p \cdot x},$$

Alternatively, one can look at the quantities of the goods purchased in the after period, x'. This is called a *Paasche Index*.

$$\text{Paasche Index } = \frac{p' \cdot x'}{p \cdot x'}$$

In practice, cost of living is computed using some variant of a Laspeyres or Paasche index. In theory, there is a better alternative, which is to use what is called an *ideal index*. Similar to the idea of EV or CV, an ideal index chooses some base level of utility, and asks how much more expensive it is to achieve this utility at prices p' than at prices p.

$$\text{Ideal Index } = \frac{e(p', u)}{e(p, u)},$$

where u is a "base" level of utility — typically either the utility in the "before" or "after" period.

Generally speaking, neither the Paasche or the Laspeyres Index are "ideal". To see why, let u be the utility in the before period. Then

$$\text{Laspeyres} = \frac{p' \cdot x}{p \cdot x} = \frac{p' \cdot x}{e(p, u)} \geq \frac{e(p', u)}{e(p, u)} = \text{Ideal}(u)$$

The problem is that at prices p', the consumer will not choose to consume x. Most likely, there is a cheaper way to get utility u. This is called the *substitution bias* because the Laspeyres index does not account for the fact that when prices change, consumers will substitute to cheaper products.

The Paasche index also suffers from substitution bias:

$$\text{Paasche} = \frac{p' \cdot x'}{p \cdot x'} = \frac{e(p', u')}{p \cdot x'} \leq \frac{e(p', u')}{e(p, u')} = \text{Ideal}(u').$$

In the last decade, the Bureau of Labor Statistics measure of inflation (the Consumer Price Index or CPI) has come under criticism. One of the main criticisms is that it suffers from substitution bias. There is also concern over the CPI for several other reasons, which include the following biases.

- “New Good Bias”. When new products are introduced, we only have a price in the after period, but not in the before period (or if products disappear, we have the opposite problem). The CPI deals with this by waiting 5–10 years to add these products (for example, cellular telephones, rice krispies treats cereal) to the index. But these products make us better off, meaning that the CPI tends to underestimate how much better off we really are. A substantial body of recent research is focused on measuring the welfare impact of new goods.

- ”Outlet Bias”. The BLS goes around and measures prices in various places, then takes an average. Over the last 20 years, people have started buying things cheaply at places like Walmart and Costco. Thus, the BLS may tend to over-estimate the prices people actually pay.

Besides price indexes for all goods, it is sometimes useful to construct price indexes for categories of goods based solely on the prices of the goods in that category. For example, one might hope to be precise about statements like "entertainment goods have become 10% more expensive" without having to refer to non-entertainment goods like food, housing, and transportation.

Ideally, we would like our price index to stand in for more detailed information in various calculations and empirical studies, especially calculations about consumer welfare and demand studies. With those intuitive goals in mind, we turn to a formal treatment.

We divide the goods $1, ..., n$ be divided into two groups. Let goods $1, ..., k$ be the ones in the category of interest, which here we call "entertainment goods," while goods $k+1, ..., n$ denote the other, non-entertainment goods. We make two assumptions and impose three requirements. The assumptions are (1) that consumer preferences are locally non-satiated and (2) that there exist some prices at which the consumer prefers to make a positive expenditure on entertainment goods. The second assumption rules out trivial cases.

The first requirement is that the entertainment price index should depend only on the prices $(p_1, ..., p_k)$ of the entertainment goods and should be homogeneous of degree 1, so that doubling all the prices doubles the index. The second requirement is that if the prices of entertainment goods change in a way that leaves the index unchanged and if the consumer's income and the prices of non-entertainment goods remain unchanged, then consumer welfare should also remain unchanged. We formulate this as the requirement that the two conditions (1) $P(p) = P(p')$ and (2) $p_j = p'_j$ for $j = k+1, ..., n$ imply that for all u, $e(p; u) = e(p', u)$. Third, one should be able to compute the demands for non-entertainment goods from the prices for those goods and the price index for entertainment goods. This is formalized by the requirement that conditions (1) and (2) above should also imply $h_j(p) = h_j(p')$ for $j = k+1, ..., n$.

The second requirement above is a separability requirement, reminiscent of the one we analyzed in the note on choice theory. In choice theory, separability was used to decompose *choices*: we required that the decision maker's ranking of choices from one set does not depend on the choices specified from another set. Here, separability is used to decompose the price vector: we require that the welfare ranking of entertainment price vectors should not depend on non-entertainment prices. As in choice theory, this separability implies a particular structure for the ranking function–there a utility function, here the expenditure function. Separability implies that there exist two functions $P : \mathbb{R}^k \to \mathbb{R}$ and $\widehat{e} : \mathbb{R}^{2+n-k} \to \mathbb{R}$, with $\widehat{e}$ increasing in its first argument, such that for all p and u,

$$e(p, u) = \widehat{e}(P(p), p_{k+1}, ..., p_n, u).$$

We leave the proof as an exercise. By Shepard's lemma, when separability applies, we may further conclude that for $j = k+1, ..., n$,

$$h_j(p,u) = \frac{\partial}{\partial p_j} e(p,u) = \frac{\partial}{\partial p_j} \widehat{e}(P(p), p_{k+1}, ..., p_n, u) \equiv \widehat{h}_j(P(p), p_{k+1}, ..., p_n, u).$$

where the last equality defines the function $\widehat{h}_j$.

Notice that the isoquants of P are the same as the "restricted" isoquants of e, where the restriction is to changes in the prices of the first k goods. If P is differentiable with non-zero derivatives, then we can characterize the slopes of the isoquants of P as follows. Applying the chain rule for any $1 \le i < j \le k$,

$$\frac{\partial e/\partial p_i}{\partial e/\partial p_j} = \frac{\partial \widehat{e}/\partial P}{\partial \widehat{e}/\partial P} \cdot \frac{\partial P/\partial p_i}{\partial P/\partial p_j} = \frac{\partial P/\partial p_i}{\partial P/\partial p_j}.$$

The construction so far guarantees the existence of a function P, but not one that is homogeneous of degree one. The next step is to convert the function P into an index $\widehat{P}$ that is homogeneous of degree 1. To that end, fix an arbitrary positive price vector prices $p > 0$ and normalize the index by setting $\widehat{P}(p) = 100$. For any price vector $p' \gg 0$, there is a unique $\alpha > 0$ such that $P(p') = P(\alpha p)$: define $\widehat{P}(p') = 100\alpha$. The index $\widehat{P}$ defined in this way is homogeneous of degree one. (We leave it as an exercise to derive the existence of such a unique α and the homogeneity of $\widehat{P}$ using our assumptions and the properties of the expenditure function.)

To summarize, a price index satisfying all the requirements set out above exists if and only if the expenditure function is separable, that is, if and only if there exist functions $\widehat{e}$ and P as described above. The separability of the expenditure function described here is different from the separability of the consumer's utility function. Neither separability condition implies the other, and both conditions can be useful for creating tractable models for both theoretical and empirical inquiries.

Part III

Producer Theory

Chapter 11

Technology

In the standard model, the firm is an entity, just like the consumer. The firm has an objective function, profit, that it wants to maximize, just as the consumer has a utility function that it wants to maximize. The consumer faces a budget constraint while the firm faces a constraint imposed by its technological capabilities. Because the firm's objectives are very similar to a consumer's objectives, many of the methods and results from consumer theory will apply to producer theory.

11.1 Features of the Standard Model

The standard model has the following features. Firms are described by fixed and exogenously given technologies that allow them to convert inputs (in simple models, these are land, labor, capital and raw materials) into outputs (products). "Competitive" producers take both input and output prices as given, and choose a production plan (a technologically feasible set of inputs and outputs) to maximize profits.

Before we get into the details, let's remark on a few key features of the model.

Price Takers

1. Firms are price takers. This competitive firm assumption applies to both input and output markets and makes it reasonable to ask questions about (1) what happens to the firm's choices when a price changes and (2) what can be inferred about a firm's technology from its choices at various price levels. For output markets, the assumption fits best when each firm has many competitors who produce perfectly substitutable products, and a parallel condition applies to input markets. Of course, even the most casual empiricism suggests that many firms sell differentiated products and have at least some flexibility in setting prices, and even small firms may have market power in buying local inputs, such as hiring workers who live near a mine or factory, so the results of the theory need to be applied with care. Even so, the pattern of analysis established in this way is often partially extendable to situations in which firms are not price takers.

Exogenous Technology

2. Technology is exogenously given. This assumption is sometimes criticized as too narrow to be useful in a world of technical change, product innovations, and consumer marketing, but it is more flexible and encompassing than most critics acknowledge. The exogenous technology model formally includes the possibility of investing in technical change, provided these investments are themselves treated as inputs into a production process. Similarly, the model formally includes advertising and branding that alter consumer's perceptions, provided that we represent these activities as transforming the output into a different product. It allows managerial effort and talent to be inputs as well, if they, too, are treated as simple inputs into production.

Profit Maximization

3. The firm maximizes profits. Since the time of Adam Smith, if not earlier, many observers have emphasized that corporations are characterized by a separation between ownership (the stockholders) and control (management), and that this separation weakens the incentives

of managers to maximize profits. The problem of motivating managers to act on behalf of owners has been a main concern for the economics (and law) of agency theory.

Students sometimes wonder about the role of assumptions such as these, particularly when they are contrary to the facts of the situation. Economists have taken a range of positions concerning how to think about simplifying assumptions, and there is no consensus about the "correct" view. One extreme position is to deny the relevance of any inference based on such models, because the premises of the model are false. At the opposite extreme, some practicing economists seem willing to accept "standard" or "customary" assumptions uncritically. Both of these extreme positions are rejected by thoughtful people.

All economic modeling abstracts from reality by making simplifying but untrue assumptions. Experience in economics and other fields shows that such assumptions models can serve useful purposes. One purpose is to support tractable models that isolate and highlight important effects for analysis by suppressing other effects. Another purpose is to serve as a basis for numerical calculations, possibly for use in estimating magnitudes, deciding economic policies, or designing economic institutions. For example, one might want to estimate the effect of a tax policy change on overall investment or hiring. The initial calculations based on a simplified model might then be adjusted to account for the effects suppressed in the model.

For a model to serve these practical purposes, its relevant predictions must be reasonably accurate. The accuracy of predictions can sometimes be checked by testing using data. Sometimes, the "robustness" of predictions can be evaluated partly by theoretical analyses. In no case, however, should models or assumptions be regarded as adequate merely because they are "usual" or "standard." Although this seems to be an obvious point, it needs to be emphasized because the temptation to skip the validation step can be a powerful one. Standard assumptions often make the theory fall into easy, recognizable patterns, while checking the suitability of the assumptions can be much harder. The validation step is not dispensable.

11.2 Production Sets

We start by describing the technological possibilities of the firm. Suppose there are n commodities in the economy. A *production plan* is a vector $y = (y_1, ..., y_n) \in \mathbb{R}^n$, where an output will have $y_k > 0$ and an input will have $y_k < 0$. If the firm has nothing to do with good k, then $y_k = 0$. The production possibilities of the firm are described by a set $Y \subset \mathbb{R}^n$, where any $y \in Y$ is feasible production plan. Figure 11.1 illustrates a production possibility set.

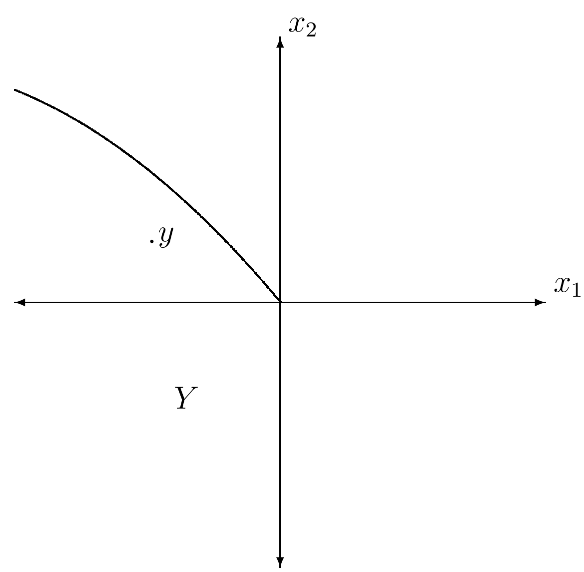

Figure 11.1: A Production Possibility Set

Throughout our analysis, we will make the innocent technical assumptions that Y is non-empty (so as to have something to study!) and closed (to make it more likely that optimal production plans exist). Consider the more interesting and substantive economic properties production sets might have:

- *Free Disposal*. The production set Y satisfies free disposal if $y \in Y$ implies that $y' \in Y$ for any $y' \leq y$.
- *Shut Down*. The production set Y has the shut-down property if $0 \in Y$; that is, the firm has the option of using no resources and producing nothing.
- *Nonincreasing Returns to Scale.* The production set Y has nonincreasing returns to scale (loosely, "decreasing returns to scale") if $y \in Y$ implies that $\alpha y \in Y$ for all $0 \leq \alpha \leq 1$.

Returns to Scale

- *Nondecreasing Returns to Scale.* The production set Y has nondecreasing returns to scale (loosely, "increasing returns to scale") if $y \in Y$ implies that $\alpha y \in Y$ for ally $\alpha \geq 1$.
- *Constant Returns to Scale.* The production set Y has constant returns to scale if $y \in Y$ implies that $\alpha y \in Y$ for all $\alpha \geq 0$.
- *Convexity.* The production set Y is convex if... Y is convex. This condition incorporates a kind of "nonincreasing returns to specialization", meaning that if two "extreme" plans are feasible, their combination will be as well. In addition, if $0 \in Y$, then convexity implies nonincreasing returns to scale.

Another way to represent production possibility sets is using a *transformation function* $T : \mathbb{R}^n \to \mathbb{R}$, where $T(y) \leq 0$ implies that y is feasible, and $T(y) > 0$ implies that y is infeasible. This is represented in Figure 11.2. You can think of the transformation function simply as a convenient way to represent a set. The set of boundary points $\{y \in \mathbb{R}^n : T(y) = 0\}$ is called the *transformation frontier*.[1]

[1]Several interpretations can be offer of the function $T(y)$. As just one example among

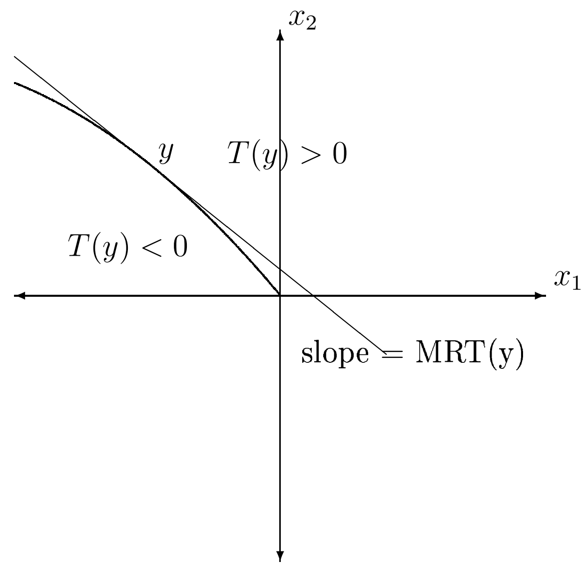

Figure 11.2: Marginal Rate of Transformation

When the transformation function is differentiable, we can define the *marginal rate of transformation of good l for good k* as:

$$MRT_{l,k}(y) = -\frac{\partial T(y)/\partial y_l}{\partial T(y)/\partial y_k}.$$

The marginal rate of transformation measures the extra amount of good k that can be obtained per unit reduction of good l. As Figure 11.2 shows, it is equal to the slope of the boundary of the production set at point y.

Thinking in terms of production sets leads to a very general model where each good k can be *either* an input or an output — that is, a firm may both produce widgets, and also use widgets to make gadgets, with y_k being the net amount of widgets produced. Often, it is convenient to separate inputs and outputs, letting $q = (q_1, ..., q_L)$ denote the vector of the firm's outputs,

many, one might interpret it to define the amount of technical progress required to make the combination y a feasible one. With that interpretation, using the currently available technology, one can produce any element of the set $\{y|T(y) \leq 0\}$.

and $z = (z_1, ..., z_M)$ the vector of inputs (where $L + M = N$).

If the firm has only a single output, we can write output as a function of the inputs used, $q = f(z)$. In this case, we refer to $f(\cdot)$ as the firm's *production function*. We can also define the *marginal rate of technological substitution* to be:

$$MRTS_{k,l}(y) = -\frac{\partial f(z)/\partial z_l}{\partial f(z)/\partial z_k}$$

The marginal rate of technological substitution tells us how much of input k must be used in place of one unit of input l to maintain the same level of output. It is illustrated in Figure 11.3

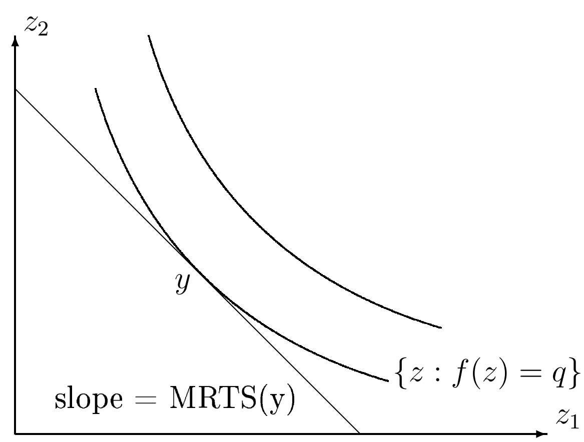

Figure 11.3: Isoquants and MRTS.

Chapter 12

Profit Maximization

In this chapter, we will study the market behavior of the firm. We will study both a maximization and a minimization problem, just as we did during consumer theory. We will assume throughout this chapter that Y is nonempty, closed, and satisfies free disposal.

12.1 The Profit Maximization Problem

We write the profit maximization problem for the firm as:

$$\begin{aligned} &\max_y p \cdot y \\ &\text{s.t. } \; y \in Y \end{aligned}$$

where we assume that $p \gg 0$.[1] The constraint that $y \in Y$ can be alternatively written as $T(y) \leq 0$.

[1] We have not yet made sufficient assumptions to ensure that a maximum exists, so it would be more proper to write $\sup_{y \in Y} p \cdot y$. The focus of our investigation here, however, will be on the properties of the maximum when it exists, rather than on conditions for a maximum to exist, so we will condition to use the "max" notation.

There are two functions of special interest in studying the problem. The first is called the "optimal production correspondence" and is denoted by $y(p)$. The correspondence $y : \mathbb{R}^n_+ \rightrightarrows Y$ maps a vector of prices into the set of profit-maximizing production plans. The second, called the "profit function," identifies the maximal value of the problem and is denoted by $\pi(p)$. That is, the profit function $\pi : \mathbb{R}^n_+ \to \mathbb{R}$ is defined to be:

$$\pi(p) = \max_{y \in Y} p \cdot y.$$

We now record some useful properties of the profit function and the optimal production correspondence. Recall that the vector notation $p > p'$ is defined by the conjunction: $p \geq p'$ and $p \neq p'$.

Properties of $\pi(p)$

PROPOSITION 1: Properties of π *The profit function π has the following properties:*

1. *$\pi(\cdot)$ is homogeneous of degree one, i.e. for all $\lambda > 0$, $\pi(\lambda p) = \lambda \pi(p)$.*
2. *$\pi(\cdot)$ is convex in p.*
3. *If Y is closed and convex, then $Y = \{y \in \mathbb{R}^N : p \cdot y \leq \pi(p)$ for all $p \in \mathbb{R}^N\}$.*
4. *If Y is closed and convex and has the free disposal property, then $Y = \{y \in \mathbb{R}^N : p \cdot y \leq \pi(p)$ for all $p \in \mathbb{R}^N_+\}$.*

Proof. (1) Note that $\pi(\lambda p) = \max_{y \in Y} \lambda p \cdot y = \lambda \max_{y \in Y} p \cdot y = \lambda \pi(p)$. □

(2) Fix p, p' and define $p^t = tp + (1-t)p'$ for $t \in [0,1]$. And let $y^t \in y(p^t)$. Then

$$t\pi(p) + (1-t)\pi(p') \geq tp \cdot y^t + (1-t)p' \cdot y^t = p^t \cdot y^t = \pi(p^t). \ \square$$

For the proof of (3) and (4) we will need the separating hyperplane theorem:

Separating Hyperplane Theorem *Let $Y \subset \mathbb{R}^N$ be a closed and convex set and suppose $x \notin Y$. Then there exists $p \in R^N$ with $p \neq 0$ such that $p \cdot x > \sup_{y \in Y}(p \cdot y)$.*

Separating Hyperplane Theorem

---

(3) Let $\widehat{Y} = \{x \in \mathbb{R}^n : p \cdot x \leq \pi(p) \text{ for all } p \in \mathbb{R}^n\}$. We need to show that $Y \subset \widehat{Y}$ and that $\widehat{Y} \subset Y$. The first inclusion follows from the definition of π. For the reverse inclusion, suppose that Y is closed and convex and $x \notin Y$. Then, by the separating hyperplane theorem, there exists $p \in \mathbb{R}^n$ such that $p \cdot x > \max_{y \in Y} p \cdot y = \pi(p)$. It follows that $x \notin \widehat{Y}$. The finding $x \notin Y \Longrightarrow x \notin \widehat{Y}$ establishes that $\widehat{Y} \subset Y$. □

(4) We argue exactly as in (3), but with one additional step. Let $\widetilde{Y} = \{y \in \mathbb{R}^n : p \cdot y \leq \pi(p) \text{ for all } p \in \mathbb{R}^N_+\}$. We need to show that $Y \subset \widetilde{Y}$ and that $\widetilde{Y} \subset Y$. The first inclusion follows from the definition of π. For the reverse inclusion, suppose that Y is closed and convex and $x \notin Y$. Then, by the separating hyperplane theorem, there exists $p \in \mathbb{R}^n$ such that $p \cdot x > \max_{y \in Y} p \cdot y = \pi(p)$. By free disposal, if any component of p were negative, then $\sup_{y \in Y} p \cdot y = +\infty$. So, no component is negative, that is, $p \in \mathbb{R}^N_+$. Therefore $x \notin \widetilde{Y}$. The finding $x \notin Y \Longrightarrow x \notin \widetilde{Y}$ establishes that $\widetilde{Y} \subset Y$. □

Q.E.D.

Proposition 2: Properties of y *The optimal production correspondence y has the following properties:*

1. *$y(\cdot)$ is homogeneous of degree zero, i.e. for all $\lambda > 0$, $y(\lambda p) = y(p)$.*
2. *If Y is convex, then for all p, the set $y(p)$ is convex. If Y is strictly convex, $p \neq 0$ and $y(p) \neq \emptyset$, then $y(p)$ is a singleton.*
3. *The Law of Supply. For any $p, p', y \in y(p)$ and $y' \in y(p')$,*

$$(p' - p)(y' - y) \geq 0.$$

Proof. (1) Note that $\pi(\lambda p) = \lambda \pi(p)$, so for $\lambda > 0$,

$$y(\lambda p) = \{y \in Y | \lambda p \cdot y = \pi(\lambda p)\} = \{y \in Y | \lambda p \cdot y = \lambda \pi(p)\} = \{y \in Y | p \cdot y = \pi(p)\} = y(p). \ \square$$

(2) Observe that $y(p) = Y \cap \{y \in \mathbb{R}^n_+ | p \cdot y = \pi(p)\}$. If Y is convex, then $y(p)$ is the intersection of two convex sets and hence is itself convex. $\square$

Suppose Y is strictly convex but $y(p)$ is not a singleton. Then for any $y \neq y' \in y(p)$, we have $y'' = \frac{1}{2}y + \frac{1}{2}y' \in interior(Y)$ and, since $y(p)$ is convex, $y'' \in y(p)$. That's impossible, because a non-trivial linear function (one with $p \neq 0$) has no local maximum. $\square$

(3) Given any $p, p', y \in y(p)$ and $y' \in y(p')$, profit maximization at price vectors p and p' imply that $p \cdot y \geq p \cdot y'$ and $p' \cdot y' \geq p' \cdot y$, respectively. So, $p \cdot (y - y') \geq 0 \geq p' \cdot (y - y')$, from which conclusion follows. $\square$

Q.E.D.

Figure 12.1: The Profit Maximization Problem

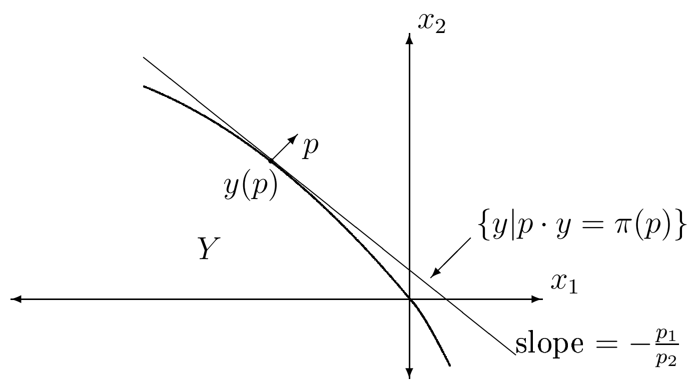

Figure 12.1 is a two-good (one output and one input) representation of the profit maximization problem.

12.2 The Envelope Theorem

Before we move on, we will formally state the envelope theorem. Consider the problem of maximizing a function $f(x, q)$, where x is a vector of endogenous variables and q is a vector of exogenous variables, under constraints:

$$\begin{aligned} &\max_{x \in \mathbb{R}^n} f(x; q) \\ &\text{s.t. } \; g(x; q) = b \end{aligned} \tag{12.1}$$

The value function of problem 12.1 is denoted $v(q)$, that is, the value attained by $f(.)$ at a solution to problem 12.1. Suppose we are interested in knowing the effect on $v(.)$ of a small change in q. Here, we will require that for values of q close to $\bar{q}$, the solution to problem 12.1 is a differentiable function $x(q)$. By the chain rule, and by noting that $v(q) = f(x(q); q)$, we have that:

$$\frac{dv(\bar{q})}{dq} = \frac{\partial f(x(\bar{q}); \bar{q})}{\partial q} + \frac{\partial f(x(\bar{q}); \bar{q})}{\partial x} \frac{dx(\bar{q})}{dq} \,. \tag{12.2}$$

But note that by the first-order conditions for unconstrained maximization that we must have

$$\frac{\partial f(x(\bar{q});\bar{q})}{\partial x} = 0\,.$$

Therefore, equation 12.2 simplifies to

$$\frac{dv(\bar{q})}{dq} = \frac{\partial f(x(\bar{q});\bar{q})}{\partial q}\,.$$

The fact that $x(q)$ is determined by maximizing the function f(.) has the implication that in computing the first-order effects of changes in q on the maximum value, we can equally well assume that the maximizer will not adjust. The only effect of any consequence is the direct effect.

Envelope Theorem

Envelope Theorem *Assume that the value function, v(q), is differentiable at $\bar{q}$ and that $(\lambda_1, \ldots, \lambda_M)$ are values of the Lagrange multipliers associated with the maximizer solution $x(\bar{q})$ at $\bar{q}$. Then*

$$\frac{dv(\bar{q})}{dq_i} = \frac{\partial f(x(\bar{q});\bar{q})}{\partial q_i} - \sum_{m=1}^{M} \lambda_m \frac{\partial g(x(\bar{q});\bar{q})}{\partial q_i}\,.$$

See MWG section M.L page 964-966 for the proof and additional detail.

12.3 Comparative Statics Results

We can now obtain two comparative statics results that relate the profit function to optimal production choices.

Proposition 3: Comparative Statics *Assume that Y is closed and satisfies free disposal. Then,*

1. ***Hotelling's Lemma****: If y is singleton-valued in a neighborhood of p, then $\pi(\cdot)$ is differentiable at p and:*

$$\frac{\partial \pi(p)}{\partial p_i} = y_i(p).$$

2. *If $y(\cdot)$ is singleton-valued and continuously differentiable, the matrix $D_p y(p) = D_p^2 \pi(p)$ is symmetric and positive semi-definite, with $[D_p y(p)]p = 0$.*

Proof. (1) Notice that the condition that y is single-valued in a neighborhood of p implies the conditions of the envelope theorem. Remember that $\pi(p) = \max_{y \in Y} p \cdot y$. Totally differentiation tells us that profits change both because price changes and because the price change changes the optimal y (the vector of inputs and outputs):

$$\frac{d\pi}{dp} = \frac{\partial \pi}{\partial p} + \frac{\partial \pi}{\partial y}\frac{\partial y}{\partial p} = y(p) + \frac{\partial \pi}{\partial y}\frac{\partial y}{\partial p}.$$

But note that the first order conditions of the profit maximization problem are $\frac{\partial \pi}{\partial y} = 0$. So if we are at an optimal y, the change in profits given by a change in p will just equal the vector $y(p)$. Therefore, it follows immediately that $\partial \pi / \partial p_i = y_i(p)$. $\square$

The first part of (2), namely that $D_p^2 \pi(p)$ is symmetric and positive semi-definite, follows from the convexity of π. For the second part of (2), observe that since $y(p)$ solves $\max_{y \in Y} p \cdot y$, it follows that p solves $\max_{p' \in \mathbb{R}^n} p \cdot y(p')$. The first-order optimality condition for that latter problem is $[D_p y(p)]p = 0$. □

Q.E.D.

Hotelling's lemma allows us to recover the firm's choices from the profit function. The symmetry of the matrix $D_p y(p)$ is a subtle empirical implication of optimization theory that was missed by economists working in a verbal tradition. Historically, this conclusion was argued to be important evidence that a mathematical approach to economic theory could lead to new insights that would be missed by a merely verbal approach.

Chapter 13

Cost Minimization with a Single Output

13.1 Conditional Factor Demand and the Cost Function

Suppose that the firm produces a single output whose quantity is denoted by q. This firm uses inputs z and faces input prices w. Using the production function notation, the firm's cost minimization problem can be written as:

$$\begin{aligned} &\min_{z \in \mathbb{R}^n_+} w \cdot z \\ &\text{s.t.} f(z) \geq q \end{aligned}$$

In our analysis of this problem, we will always assume that all input prices are strictly positive: $w \gg 0$. As usual in our study of optimization problems, two functions are of central interest. The first is the solution to the problem, $z(q, w)$. We refer to $z(q, w)$ as the *conditional factor demand* to indicate that it is conditional on a fixed level of output q. The second is the optimal value function. The optimal value function for this problem is:

$$c(q, w) = \min_{\{z: f(z) \geq q\}} w \cdot z,$$

It is called the *cost function* gives the minimum cost at which output q can be produced.

If $f(z)$ is differentiable and concave, we can use the Kuhn-Tucker method to solve for the conditional factor demands. The Lagrangian for the cost problem is:

$$\max_{\lambda \geq 0, \mu \geq 0} \min_{z} w \cdot z - \lambda \left[f(z) - q \right] - \sum_{i=1}^{m} \mu_i z_i.$$

The first-order conditions from the Lagrangian problem are:

$$\lambda \frac{\partial f(z)}{\partial z_i} \leq w_i \qquad \text{with equality if } z_i > 0.$$

and of course the solution must satisfy the production constraint. $f(z) \geq q$.

Later, we will compare these first-order conditions to the ones arising from the profit maximization problem. In the meantime, we record a few properties of the cost function. (For a full recital, including properties of the conditional factor demands, see MWG, Proposition 5.C.2.)

Proposition 4: Properties of $c(.)$ *The cost function c has the following properties:*

1. $c(\cdot)$ *is homogeneous of degree one in* w
2. $c(.)$ *is increasing in* q.
3. $c(\cdot)$ *is a concave function of* w.
4. *If* $f(\cdot)$ *is concave, then* $c(\cdot)$ *is a convex function of* q *(i.e. marginal costs are increasing in* q*).*

5. *Shepard's Lemma: If $z(\cdot)$ is single-valued, then $c(\cdot)$ is differentiable with respect to w and*

$$\frac{\partial c(q,w)}{\partial w_i} = z_i(q,w)$$

6. *If $z(\cdot)$ is a differentiable function, then the matrix $D_w z(q,w) = D_w^2 c(q,w)$ is symmetric and negative semi-definite, and $D_w z(w,q)w = 0$.*

Proof. (1) The cost function $c(q,w)$ is homogeneous of degree one in w: $c(q,\lambda w) = \min_{z:f(z)\geq q} \lambda w \cdot z = \lambda \min_{z:f(z)\geq q} w \cdot z = \lambda c(q,w)$.□

(2) The cost function $c(q,w)$ is increasing in q. To show this consider $q' > q$ and suppose that $c(q',w) < c(q,w)$. But then because we have free disposal, a less expensive way to produce q would be to produce q' and then throw away $(q'-q)$. The firm could, in this way, produce output q for a cost of $c(q',w) < c(q,w)$. But this contradicts that $c(q,w)$ is the cost at the solution to the cost minimization problem. Therefore, $c(q,w)$ is increasing in q. □

(3) Choose any two factor price vectors w and w' and some $t \in [0,1]$,and let $w'' = tw + (1-t)w'$. Let z be a solution to the firm's cost minimization problem at (q,w''), then $f(z) \geq q$. If the firm was faced with input prices w or w', z could also be used to produce q, however, at these input prices, there might be something better than z. Thus, $c(q,w) \leq w \cdot z$ and $c(q,w') \leq w' \cdot z$. Therefore, $tc(q,w)+(1-t)c(q,w') \leq tw \cdot z + (1-t)w' \cdot z = w'' \cdot z = c(q,w'')$. □

(4) Left for you to prove. □

(5) We know that

$$c(q,w) = w \cdot z(q,w)\,.$$

Differentiating both sides with respect to w_i gives:

$$\frac{\partial c(q,w)}{\partial w_i} = z_i(q,w) + \sum_j w_j \frac{\partial z_j(q,w)}{w_i}\,.$$

What this means is that if we raise the price of input i, the resulting change in cost needed to produce output level q comes from two terms. First, the amount z_i that the firm demanded before is now more expensive and second, the firm will probably want to use less of input i now that it is more expensive and will also want to change the amounts of other inputs demanded (the firm will want to use more of inputs that are substitutes for input i and will want to use less of inputs that are complements of input i). The result that we want is that this summation term is zero. Recall that $z(q,w)$ is the solution of the cost minimization problem. The first-order condition for z_i is

$$w_i = \lambda \frac{\partial f(z)}{z_i}\,.$$

This still isn't quite what we want, so consider the identity $f(z(q,w)) = q$. If we differentiate both sides with respect to w_i we get

$$\sum_j \frac{\partial f}{\partial z_j} \frac{\partial z(q,w)_j}{\partial w_i} = 0\,.$$

If we insert the first-order condition, we obtain

$$\frac{1}{\lambda} \sum_j w_j \frac{\partial z(q,w)_j}{\partial w_i} = 0\,.$$

This is what we want. As long as λ isn't zero or infinite, $\sum_j w_j \frac{\partial z(q,w)_j}{\partial w_i} = 0$ and therefore,

$$\frac{\partial c(q,w)}{\partial w_i} = z_i(q,w).\ \square$$

(6) Left for you to prove. $\square$

Q.E.D.

13.2 Price Equals Marginal Cost Condition

Returning to our characterization of the firm's problem, suppose the firm solves the cost minimization problem for every q, giving it a cost function $c(q, w)$. The profit maximization problem can then be seen as:

$$\max_q pq - c(q, w).$$

This problem gives the famous first-order conditions:

$$p = \frac{\partial c(q, w)}{\partial q},$$

equating price and marginal cost. So profit maximization implies that the correct shadow price is the market price for output p.

13.3 Profit Maximization with a Single Output

Consider the profit maximization problem when there is only a single output. With a vector $z = (z_1, ..., z_M)$ of inputs and a single output $f(z)$, the profit maximization problem can be simplified to:

$$\max_{z \in \overline{\mathbb{R}}^m_+} p \cdot f(z) - w \cdot z,$$

where p reflects the price of output and the vector $w \gg 0$ reflects the input prices. Denote the maximizing solution by $z(p, w)$. There may be multiple solutions or no solution for some price vectors (p, w), but we will focus our analysis on the cases where the solution is a singleton, in which case $z(p, w)$ denotes the *factor demands* at prices (p, w). If $f(z)$ is differentiable and concave, the factor demands (when they exist!) can be found by solving the first-order conditions: for all i,

$$p\frac{\partial f(z)}{\partial z_i} \leq w_i \text{ with equality if } z_i > 0.$$

It is interesting to compare this solution to that of the cost minimization problem. There, the first order conditions were that

$$\lambda\frac{\partial f(z)}{\partial z_i} \leq w_i \text{ with equality if } z_i > 0$$

Thus, with f concave, one can think of profit maximization as the special case of cost minimization in which the shadow price of output is the market price p. There is more to this account. From the envelope theorem, we have:

$$\lambda = \frac{\partial c(q, w)}{\partial q}.$$

Thus, at the solution to the cost minimization problem, the shadow value of output λ is exactly the marginal cost of production.

When f is concave, the approach based on first-order conditions is useful for working examples and obtaining formula that can be used to compute solutions numerically. The convexity assumption fails in several interesting cases, such as ones where there are fixed costs of production or where the production sets exhibits increasing returns. It turns out that comparative statics conclusions are largely independent of convexity assumptions, so we approach the problem of comparative statics using methods that do not rely on convexity. The first result is an easy one that follows from the law of supply.

Chapter 14

Monotone Comparative Statics

This chapter examines how the firm's optimal production choice (or in the case of monopoly, price) changes for increases or decreases in exogenous parameters (usually prices). These techniques can be applied to a variety of applications, however, because the consumer's choice set is often not a lattice, these techniques are not usually applied in consumer theory.

14.1 Supermodularity and Isotone Differences

Definition: Join $z = x \vee y$ means $z_i = \max\{x_i, y_i\}\ \forall\, i = 1, \ldots, N$

Definition: Meet $z = x \wedge y$ means $z_i = \min\{x_i, y_i\}\ \forall\, i = 1, \ldots, N$

We will need to explicitly define $\geq$ on $\mathbb{R}^N$. We will say that $x \geq y$ if $\forall\, i = 1, \ldots, N$, $x_i \geq y_i$. This is a partial order, which is to say that $(\mathbb{R}^N, \geq)$ is a partially ordered set.

Definition: Isotone in Differences A function $f(x, w)$ is said to be isotone (increasing) in differences if for $x' \geq x$, $f(x', w) - f(x, w)$ is increasing in w. Isotone Differences

---

Supermodular

DEFINITION: Supermodular A function $f : \mathbb{R}^N \to \mathbb{R}$ is supermodular if $\forall\, x, y \in \mathbb{R}^N$ $f(x) + f(y) \leq f(x \wedge y) + f(x \vee y)$.

Now that we have these definitions, is is natural to consider the relationship between these two concepts.

> **Question:** Is there any difference between supermodularity and isotone differences?
>
> **Answer:** If $f : \mathbb{R}^2 \to \mathbb{R}$, then the answer is no. However, in general there is a difference. Consider the following function
>
> $$f(x, y, z) = xyz - yz$$
>
> This function satisfies isotone (increasing) differences in $(x; y, z)$, but is not supermodular in (x, y, z). Supermodularity requires isotone differences between every pair of variables. Notice that for $x < 1$, yz has the wrong sign.

DEFINITION: Lattice A partially ordered set X is a lattice if $\forall\, x, y \in X$, $x \vee y \in X$ and $x \wedge y \in X$.

Note, this is very important because a function can only be supermodular if it is defined over a lattice. This is why these techniques cannot always be applied in consumer problems; the choice set is not usually a lattice.

Consider a function $f(x; \theta)$ that depends on a vector of choice variables x and a vector of exogenous parameters θ. Table 14.1 illustrates the definitions of supermodularity and isotone (increasing) differences in a general and differentiable setting (where $f(.)$ is twice continuously differentiable).

Table 14.1: Supermodularity and Isotone Differences

Condition	**General Setting**	**Differentiable Setting**
Supermodular in x	$\forall\, x, y \in X$ and $\forall \theta \in \Theta$ $f(x \vee y; \theta) + f(x \wedge y; \theta) \geq f(x;\theta) + f(y;\theta)$	$\frac{\partial^2 f}{\partial x_i \partial x_j} \geq 0$ for $i \neq j$
Supermodular in θ	$\forall\, x \in X$ and $\forall \theta, \theta' \in \Theta$ $f(x, \theta \vee \theta') + f(x, \theta \wedge \theta') \geq f(x;\theta) + f(x;\theta')$	$\frac{\partial^2 f}{\partial \theta_i \partial \theta_j} \geq 0$ for $i \neq j$
Isotone Differences in (x, θ)	$\forall\, x' \geq x$ and $\forall \theta' \geq \theta$ $f(x', \theta') - f(x, \theta') \geq f(x', \theta) - f(x, \theta)$	$\frac{\partial^2 f}{\partial x_i \partial \theta_j} \geq 0\ \forall\, i, j$
Supermodular in (x, θ)	$\forall\, (x, \theta), (x', \theta') \in X \times \Theta$ $f((x,\theta) \vee (x', \theta')) + f((x,\theta) \wedge (x', \theta'))$ $\geq f(x;\theta) + f(x';\theta')$	All of the above

14.2 The Topkis Monotonicity Theorem

We will consider the general optimisation problem:

$$V(\theta) = \max_{x \in X} \{f(x, \theta) + g(x)\}$$

where the function $f(x, \theta)$ maps the space $X \times \Theta$ to the real numbers $\mathbb{R}$ and $g(x)$ maps X to $\mathbb{R}$. We will then define the correspondence:

$$x^*(\theta; g) = \arg\max_{x \in X} \{f(x, \theta) + g(x)\}$$

to be the set of optimal choices given θ and g. We want to examine whether $x^*(\theta; g)$ is increasing or decreasing in θ.

Topkis Monotonicity Theorem *Assume that* $X \subseteq \mathbb{R}^n$ *is a lattice and that* Θ *is a partially ordered set. If* $g(.)$ *is supermodular, and if* $f(.)$ *is supermodular in* x *and has isotone (increasing) differences in* $(x; \theta)$*, then*

$$x^*(\theta'; g) \geq x^*(\theta; g) \; \forall \, \theta' \geq \theta$$

that is, $x^*(\theta; g)$ *is nondecreasing in* θ*.*

14.2.1 A Standard Example

Consider the following maximization problem:

$$\max_{k,l \geq 0} \{pf(k,l) - wl - rk\}$$

where we assume that output is increasing in both inputs and that capital k and labor l are substitutes, that is $f_{lk} \leq 0$, so that increasing the usage of one input reduces the marginal product of the other input. Assume in the short run that capital is fixed and so the firm can only choose to vary labor. Let

$$\Omega(\bar{k}, l; p, w, r) = pf(\bar{k}, l) - wl - r\bar{k}$$

since k is fixed at $\bar{k}$, the only choice variable is l. Since there is only one control variable, $\Omega(\bar{k}, l; p, w, r)$ is supermodular in l. So, to apply the Topkis Monotonicity Theorem, we need to show isotone differences. To do this, we need to check that the cross partial derivatives between labor and each of the parameters are non-negative.

$$\frac{\partial \Omega}{\partial l} = pf_l - w$$

So now we check to see if the cross partials are non-negative:

$$\frac{\partial^2 \Omega}{\partial l \partial p} = f_l \geq 0$$

$$\frac{\partial^2 \Omega}{\partial l \partial r} = 0 \geq 0$$

$$\frac{\partial^2 \Omega}{\partial l \partial w} = -1 \not\geq 0$$

So we cannot apply the Topkis Theorem. However, if we define a new parameter $\tilde{w} = -w$, then we can re-write the problem as:

$$\Omega(\bar{k}, l; p, \tilde{w}, r) = pf(\bar{k}, l) + \tilde{w}l - r\bar{k}$$

Now we have that

$$\frac{\partial^2 \Omega}{\partial l \partial \tilde{w}} = 1 \geq 0$$

So we have shown that $\Omega(\bar{k}, l; p, \tilde{w}, r)$ is supermodular in l and has increasing differences in $(l; p)$, $(l; r)$, and $(l; \tilde{w})$. Thus, by the Topkis Monotonicity Theorem, $l(p, \tilde{w}, r, \bar{k})$ is nondecreasing in p, r, and $\tilde{w}$. Which means that $l(p, \tilde{w}, r, \bar{k})$ is nonincreasing in $\tilde{w}$.

14.2.2 Conditional Factor Demand and Prices

This last point in the example, that $l(p, \tilde{w}, r, \bar{k})$ is nonincreasing in $\tilde{w}$, can be shown to hold generally. Here we apply monotone comparative statics to prove a result about how conditional factor demands will change with prices.

Proposition 5: $\forall\ i,\ q,$ $\;z_i(q, w)$ *is nonincreasing in* w_i.

This proposition simply states that conditional factor demands are weakly decreasing in own price. That is, if the price of an input goes up, the firm's demand for that input goes down.

Proof. The traditional proof, and still the easiest, mimics the argument for the law of supply. But rather than take the easy way out, this is a good chance to illustrate the Topkis Theorem. Consider the firm's cost minimization problem

$$\min_{\{z \in \mathbb{R}^n_+ | f(z) \leq q\}} w \cdot z$$

and let us consider maximizing the negative of the objective rather than minimizing the objective. Now fix z_i and let $\lambda(q, w; z_i)$, $\mu(q, w; z_i)$ and $z_{-i}(q, w; z_i)$ denote the optimal (Lagrangian minimizing) choices taking z_i as fixed. It is easy to see that changes in w_i have no effect on these choices because z_i is fixed. The problem of finding the optimal conditional factor demand z_i is then to solve:

$$\max_{z_i} -w_i \cdot z_i + \lambda(q, w_{-i}; z_i) \cdot [f(z_i, z_{-i}(q, w_{-i}; z_i)) - q] + \mu_i(q, w_{-i}; z_i) \cdot z_i$$
$$+ \text{ terms not depending on } z_i.$$

This problem has increasing differences in $(z_i, -w_i)$. Applying Topkis' Theorem, we see that $z_i(q, w)$ will be decreasing in w_i. (Question: can you identify conditions under which $z_i(q, w)$ would also decrease in w_j for $j \neq i$ — think about complements and substitutes in production?)

Q.E.D.

14.2.3 Monopoly Example

Consider a monopolist who faces a fixed marginal cost of production c and demand given by the function $D(p)$. The monopolist chooses the output price p so as to maximise $\Pi(p, c) = (p - c)D(p)$. Since the objective function is a function of only one control variable, it is trivially supermodular in p (make sure that you know how to show this). However, what about increasing differences?

$$\frac{\partial \Pi}{\partial p} = D'(p)(p-c) + D(p)$$

$$\frac{\partial^2 \Pi}{\partial p \partial c} = -D'(p)$$

So whether the objective function in this form has increasing differences in $(p;c)$ depends on the differentiability and shape of the demand function. In most cases, when it exists, $D'(p) \leq 0$, so $\frac{\partial^2 \Pi}{\partial p \partial c} \geq 0$. However, we can transform the problem so that no assumptions about the demand function are necessary. Consider taking a logarithmic transformation of the objective function. Since $\ln(.)$ is an increasing function, the argument that maximises $\Pi(p,c)$ also maximises $\ln(\Pi(p,c))$, given below:

$$\ln(\Pi(p,c)) = \ln(p-c) + \ln(D(p)).$$

Then we get

$$\frac{\partial \ln(\Pi)}{\partial p} = \frac{1}{p-c} + \frac{D'(p)}{D(p)}$$

$$\frac{\partial^2 \ln(\Pi)}{\partial p \partial c} = \frac{1}{(p-c)^2} \geq 0$$

As shown above, the objective function is supermodular in p and has increasing differences in $(p;c)$ and this result was obtained without making any assumptions about the demand function. We have shown that the conclusion that price increases (is nondecreasing) for an increase in marginal cost c holds regardless of the properties of the demand function.

Chapter 15

Complements and Substitutes

Informally, two inputs are called substitutes when an increase in the price of one leads to an increase in input demand for the second and complements when it leads to a decrease in input demand for the second. That is,

$$\text{if } \frac{\partial z_i}{\partial w_j} \geq 0 \text{ then we say that input } z_i \text{ and } z_j \text{ are complements}$$

$$\text{if } \frac{\partial z_i}{\partial w_j} \leq 0 \text{ then we say that input } z_i \text{ and } z_j \text{ are substitutes}$$

Several things conspire to complicate this seemingly simple definition.

First, it is perfectly possible that the change in demand in response to a price increase is not uniform; for example, the demand for input j may increase as the price of input i increases from w_i to w_i' and may then decrease as the input price increases further to w_i''. Inputs i and j may be substitutes at some prices and then may be complements at other prices.

Second, the response to a price increase can depend on which optimization problem we are using to determine demand. Input demands might correspond to the solution of a cost minimization problem in which output q is held constant, or from the long-run profit maximization, with all inputs and output free to vary, or a short-run profit maximization problem, with some inputs fixed.

15.1 Profit Maximization Input Substitution

Let us consider the profit maximization problem for a single-output firm with two inputs:

$$\max_z \{pf(z_1, z_2) - w \cdot z\}.$$

If $f(z)$ is differentiable, the demand function $z(p, w)$ must satisfy the necessary first-order conditions

$$\begin{aligned} p\frac{\partial f(z_1(p, w_1, w_2), z_2(p, w_1, w_2)}{\partial z_1} &\equiv w_1 \\ p\frac{\partial f(z_1(p, w_1, w_2), z_2(p, w_1, w_2)}{\partial z_2} &\equiv w_2. \end{aligned}$$

To simplify the notation, we will normalize the output price to be one ($p = 1$). Differentiating the first-order conditions with respect to w_1 we have

$$\begin{aligned} f_{11}\frac{\partial z_1}{\partial w_1} + f_{12}\frac{\partial z_2}{\partial w_1} &= 1 \\ f_{21}\frac{\partial z_1}{\partial w_1} + f_{22}\frac{\partial z_2}{\partial w_1} &= 0. \end{aligned}$$

Differentiating with respect to w_2 we have

$$\begin{aligned} f_{11}\frac{\partial z_1}{\partial w_2} + f_{12}\frac{\partial z_2}{\partial w_2} &= 0 \\ f_{21}\frac{\partial z_1}{\partial w_2} + f_{22}\frac{\partial z_2}{\partial w_2} &= 1. \end{aligned}$$

For simplicity, we will write this in matrix form

$$\begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial w_1} & \frac{\partial z_1}{\partial w_2} \\ \frac{\partial z_2}{\partial w_1} & \frac{\partial z_2}{\partial w_2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

If the Hessian matrix, $\begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix}$, is invertible, that is, if we have a regular maximum, then we can solve for the matrix of first derivatives

$$\begin{bmatrix} \frac{\partial z_1(p,w)}{\partial w_1} & \frac{\partial z_1(p,w)}{\partial w_2} \\ \frac{\partial z_2(p,w)}{\partial w_1} & \frac{\partial z_2(p,w)}{\partial w_2} \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix}^{-1}$$

The matrix on the left is known as the substitution matrix since it describes how the firm substitutes one input for another as the factor prices change. The second-order condition for profit maximization is that the Hessian matrix is a symmetric negative definite matrix. The inverse of a symmetric negative definite matrix is also a symmetric negative definite matrix. This means that the substitution matrix is also symmetric and negative definite (how does this relate to proposition 3?). This tells us that the firm's demand for input i when price j changes is equal to the change in the firm's demand for input j when price i changes.

15.2 Cost Minimization Input Substitution

Now consider the firm's cost minimization problem for a single-output firm with two inputs

$$\min_{\{z|f(z)\geq q\}} \{w_1 z_1 + w_2 z_2\}.$$

The first-order conditions are

$$\begin{aligned} f\left(z_1(q,w),\, z_2(q,w)\right) &\equiv q \\ \lambda \frac{\partial f\left(z_1(q,w),\, z_2(q,w)\right)}{\partial z_1} &\equiv w_1 \\ \lambda \frac{\partial f\left(z_1(q,w),\, z_2(q,w)\right)}{\partial z_2} &\equiv w_2. \end{aligned}$$

Following the same approach as with profit maximization[1] and applying Cramer's rule, we obtain the following substitution matrix

$$\begin{bmatrix} \frac{\partial z_1(q,w)}{\partial w_1} & \frac{\partial z_1(q,w)}{\partial w_2} \\ \frac{\partial z_2(q,w)}{\partial w_1} & \frac{\partial z_2(q,w)}{\partial w_2} \end{bmatrix} = \begin{bmatrix} \frac{f_2^2}{H} & \frac{-f_1 f_2}{H} \\ \frac{-f_2 f_1}{H} & \frac{f_1^2}{H} \end{bmatrix}.$$

In this two-input case, the sign of the cross price effect must be positive, that is, $\frac{-f_1 f_2}{H} > 0$.[2] The two factors must be substitutes. This is special to the

[1] Differentiating with respect to w_1 we find that

$$\begin{aligned} \frac{\partial f}{\partial z_1}\frac{\partial z_1}{\partial w_1} + \frac{\partial f}{\partial z_2}\frac{\partial z_2}{\partial w_1} &\equiv 0 \\ 1 - \lambda\left[\frac{\partial^2 f}{\partial z_1^2}\frac{\partial z_1}{\partial w_1} + \frac{\partial^2 f}{\partial z_1 \partial z_2}\frac{\partial z_2}{\partial w_1}\right] - \frac{\partial f}{\partial z_1}\frac{\partial \lambda}{\partial w_1} &\equiv 0 \\ 0 - \lambda\left[\frac{\partial^2 f}{\partial z_2 \partial z_1}\frac{\partial z_1}{\partial w_1} + \frac{\partial^2 f}{\partial z_2^2}\frac{\partial z_2}{\partial w_1}\right] - \frac{\partial f}{\partial z_2}\frac{\partial \lambda}{\partial w_1} &\equiv 0. \end{aligned}$$

Using Cramer's rule, we can solve for how the demand for input 2 changes for price 1:

$$\frac{\partial z_2(q,w)}{\partial w_1} = \frac{\det\begin{bmatrix} 0 & -f_1 & 0 \\ -f_1 & -\lambda f_{11} & -1 \\ -f_2 & -\lambda f_{12} & 0 \end{bmatrix}}{\det\begin{bmatrix} 0 & -f_1 & -f_2 \\ -f_1 & -\lambda f_{11} & -\lambda f_{21} \\ -f_2 & -\lambda f_{12} & -\lambda f_{22} \end{bmatrix}} > 0$$

[2]

$$\text{where } H = \det\begin{bmatrix} 0 & -f_1 & -f_2 \\ -f_1 & -\lambda f_{11} & -\lambda f_{21} \\ -f_2 & -\lambda f_{12} & -\lambda f_{22} \end{bmatrix}$$

two-input case; if there are more factors of production, the cross-price effect between any two of them can be positive or negative.

Question: Why is it that in the two-input cost minimization problem, the inputs must be substitutes? Why do we not have this condition in the two-input profit maximization problem?

15.3 Supermodularity and Input Substitution

However, in producer theory, there are perhaps even better[3] definitions of substitutes and complements. Rather than focus on the choice behavior, that is, how the firm's choice of inputs z changes for a given input price change, we could focus on the production function itself. Consider the question of how does the marginal productivity of input i change as we change the amount of input j:

$$\frac{\partial}{z_j}\left(\frac{\partial f}{\partial z_i}\right) = \frac{\partial^2 f}{\partial z_i \partial z_j} = f_{ij}$$

If $f_{ij} > 0$, then inputs i and j are complements, that is, if the firm increases the amount of input j that it uses, the marginal product of input i increases and therefore the firm will demand more of input i. If $f_{ij} < 0$, then inputs i and j are substitutes, that is, if the firm increases the amount of input j that it uses, the marginal product of input i decreases and therefore the firm will demand less of input i.

These definitions of input substitution are closely related to supermodularity. If $f(z)$ is differentiable, then saying that $f(z)$ is supermodular in z is the same as saying that $f_{ij} \geq 0$ for all inputs $i \neq j$. This notion of complements implies that the inputs are complements in the profit maximization problem. Proposition 6 shows this relationship for the profit maximization problem

[3]Better in the sense that here, the definition is based on the primitives (the production function) and not on the firm's behavior

with a single kind of output and all inputs free to vary. It tells us that, under certain conditions, if $f(z)$ is supermodular then $\frac{\partial z_i(p,w)}{\partial w_j} \leq 0$ for all i and j.

Proposition 6 *Restrict attention to the domain of price vectors $(p, w) \in \mathbb{R}_+^{n+1}$ upon which $z(p, w)$ is singleton-valued. If $f(z)$ is increasing and supermodular ($f_{ij}(z) \geq 0$), then $z(p, w)$ is isotone ("weakly increasing") in p and antitone ("weakly decreasing") in w.*

Proof. Since $f(\cdot)$ is increasing and supermodular, the firm's objective function $p \cdot f(z) - w \cdot z$ is supermodular in (z, p). Also, the choice set $\mathbb{R}_+^n$ is a sublattice. So by Topkis' Monotonicity Theorem, $z(p, w)$ must be increasing in p. Similarly, the firm's objective is also supermodular in in $(z, -w_i)$. So $z(p, w)$ is antitone in w_i. □

Q.E.D.

The preceding proposition is easily extended to "short run" profit maximization. Consider the problem in which some set of inputs S is held fixed at the levels x_S. Define $z(p, w, x_S)$ to be the solution to the firm's profit maximization problem given the extra constraint $z_S = x_S$. This additional constraint defines a *sublattice*, so the original proof still applies.

Supermodularity is however stronger than the long-run price-theory notion of complementary, because it implies the price theory concept not only for the long-run problem but also for all possible short-run problems. It is stronger in another way, as well: it characterizes the behavior of f even around choices z that would never be justified by any price vector.

The next proposition asserts that when f is strictly concave, so each choice is the unique optimum for some set of prices, then supermodularity is identical to this profit maximization notion of complementarity.

Proposition 7 *Suppose that f is increasing and strictly concave. If for all S and x_S, $z(p, w, x_S)$ is antitone in w, then f is supermodular.*

Proof. Left as an exercise. (Hint: Suppose that all but 2 inputs are fixed. How does f vary in the remaining two inputs?)

Q.E.D

Note that by Topkis's Monotonicity Theorem, if f is supermodular, then $z(p, w, x_S)$ is antitone in w, for all S, even without the assumptions that f is increasing and concave. So, the import of the proposition is that for the case of an increasing concave production function f, inputs are complements in the strong sense defined by the proposition if and only if f.is supermodular.

The substitutes case is similar for the two-input case, but subtler for the general cases. Proposition 8 shows the relationship between supermodularity and substitutes for the two-input case. Proposition 9 gives the relationship between submodularity and complements for the multiple-input case. Note that Proposition 8 concerns $z(p, w)$, the solution to the single-output profit maximization problem. Because there are only two inputs, $z_1(q, w)$ and $z_2(q, w)$ are substitutes regardless of our assumptions on $f(z)$.

Proposition 8 *Restrict attention to the domain of price vectors (p, w) upon which $z(p, w)$ is singleton-valued and suppose there are just two inputs. If $-f(z)$ is supermodular, then $z_1(p, w)$ is everywhere non-decreasing in w_2 and $z_2(p, w)$ is everywhere non-decreasing in w_1.*

Proof. With two inputs, define $\widehat{f}(z_1, -z_2) = f(z)$. If $-f$ is supermodular, then the function $\widehat{f}$ is also supermodular. The conclusion then follows directly by applying Topkis's Monotonicity Theorem to the problem: $\max_{x_1 \geq 0, x_2 \leq 0} p\widehat{f}(x) - w_1 x_1 + w_2 x_2$. □

Q.E.D.

For the case of multiple inputs, one can obtain characterizations of complements using the indirect profit function or the cost function, as follows.

Proposition 9 *Suppose that the cost function $c(q, w)$ is continuously differentiable. Then, the following two are equivalent: (1) for all $i \neq j$ and all w_{-j} and q, $z_i(q, w)$ is non-increasing in w_j, and (2) for all q, $c(q, w)$ is submodular in $-w$.*

Proof. By Shepard's lemma, $z_i(q, w) = \frac{\partial}{\partial w_i} c(q, w)$. Consider differentiating both sides with respect to w_j.

$$\frac{\partial z_i(q, w)}{\partial w_j} = \frac{\partial^2 c(q, w)}{\partial w_i \partial w_j}$$

From here is is clear that $z_i(q, w)$ is always non-increasing in w_j if and only if $\frac{\partial^2 c(q,w)}{\partial w_i \partial w_j} \leq 0$, that is if $c(q, w)$ is submodular in w. □

Q.E.D.

Proposition 10 is a parallel proposition for $z(p, w)$, the solution to the firm's profit maximization problem. It is proved similarly, using Hotelling's lemma in place of Shepard's lemma.

Proposition 10 *Suppose that the profit function $\pi(p, w)$ is continuously differentiable. Then, the following two are equivalent: (1) for all $i \neq j$ and all w_{-j} and p, $z_i(p, w)$ is non-increasing in w_j, and (2) for all p, $\pi(p, w)$ is supermodular in w.*

Proof. By Hotelling's Lemma, $-z_i(p,w) = \frac{\partial}{\partial w_i}\pi(p,w)$. Consider differentiating both sides with respect to w_j.

$$\frac{\partial z_i(p,w)}{\partial w_j} = -\frac{\partial^2 \pi(p,w)}{\partial w_i \partial w_j}$$

From here is is clear that $z_i(p,w)$ is always non-increasing in w_j if and only if $\frac{\partial^2 \pi(p,w)}{\partial w_i \partial w_j} \geq 0$, that is if $\pi(p,w)$ is supermodular in w. □

Q.E.D.

Chapter 16

The Short-Run and Long-Run

While not treating time explicitly, the neoclassical theory of the firm typically distinguishes between the *long-run*, a length of time over which the firm has the opportunity to adjust all factors of production, and the *short-run*, during which time some factors may be difficult or impossible to adjust.

In his *Foundations of Economic Analysis* (1947), Samuelson suggested that a firm would react more to input price changes in the long-run than in the short-run, because it has more inputs that it can adjust. This view still persists in some economics texts.[1] Samuelson called this effect the *LeChatelier principle* and argued that it also illuminates how war-time rationing makes demand for non-rationed goods less elastic. Assuming that the optimal production choice $y(p)$ is differentiable, he proved that the principle holds for sufficiently small price changes in a neighborhood of the long-run price. The relation between long and short run effects can be quite important, because data about the short-run effects of policies are frequently used to forecast their long-run effects, and such forecasts can influence policymakers.

[1] For example, Varian (1992) writes: "It seems plausible that the firm will respond more to a price change in the long run since, by definition, it has more factors to adjust in the long run than in the short run. This intuitive proposition can be proved rigorously."

We begin our analysis with an example to prove that the Samuelson-LeChatelier principle does not apply to large price changes. Consider the production set $Y = \{(0,0,0), (1,-2,0), (1,-1,-1)\}$ in which two inputs, goods 2 and 3, are used to produce good 1. The set includes the possibilities that the firm can produce a unit of output either by using two units of good 2 or by using one unit of each kind of input. The third and last possibility is that the firm can use no inputs and produce nothing.

Suppose long-run prices are given initially by $p = (2, .7, .8)$. At the corresponding initial long-run optimum, the firm achieves its maximum profit of 0.6 by choosing the point $x^{LR}(p) = (1,-2,0) \in Y$. The superscript designates this as a *long-run* choice–one that treats all points in the set Y as feasible. Suppose that the use of good 3 is fixed in the short run and that the price of good 2 rises to 1.1, so the new price vector is $p' = (2, 1.1, .8)$. Since the firm cannot immediately change its use of good 3, it must choose between its current plan $(1,-2,0)$ incurring a loss $(profit = -0.2)$ or switching to $(0,0,0)$ $(profit = 0)$. The latter choice maximizes its profit, so $x^{SR}(p', p) = (0,0,0)$, where the notation indicates that this is the firm's profit-maximizing *short-run* choice when current prices are p' but fixed inputs were chosen when prices were p. The firm's long run choice at price vector p' is $x^{LR}(p') = (1,-1,-1)$. In this example, when the price of the first input rises, the demand for good 2 changes in the short-run from $x_2^{LR}(p) = -2$ to $x_2^{SR}(p', p) = 0$, but then recovers in the long-run to $x_2^{LR}(p') = -1$. So, the short-run change is *larger* than the long-run change, contrary to the Samuelsonian conclusion.

Although the three-point production set may seem unusual, the example can be modified to make Y convex and smooth. The first step is to replace Y by its convex hull $\widehat{Y}$ (the triangle with vertexes at the three points in the original set Y). The choices from $\widehat{Y}$ are the same as those from Y, so that gives us a convex model of the same choices. One can further expand $\widehat{Y}$ by adding free disposal without changing the preceding calculations. For a similar example with a strictly convex production set having a smoothly curved boundary, one can replace $\widehat{Y}$ by the set $\widehat{Y}_\epsilon \subset \{y \in \mathbb{R}^3_+ | (\exists x \in \widehat{Y}) |y-x| \leq \varepsilon\}$, where $\varepsilon > 0$. For ε small, the firm's profit-maximizing choices from $\widehat{Y}_\epsilon$ differ little from its profit-maximizing choices from Y. In particular, the long-run response to the price change will remain smaller than the short-run response.

There is an interesting set of economic models in which it is always true that long-run responses to price changes are larger than short run responses. Intuitively, these are models in which a "positive feedbacks" argument applies, as follows.

Suppose that output is not fixed and that there are two inputs, capital and labor, which are *substitutes*. Suppose that capital is fixed in the short-run. By the law of demand, if the wage increases, the firm will use less labor both in the short-run and in the long-run. Since the two inputs are substitutes, the increased wage implies an increased use of capital in the long-run. Since $f_{kl} \leq 0$ for substitutes, the additional capital used in the long-run will reduce the marginal product of labor, so in the long-run the firm will use still less labor. In summary, the long-run effect is larger than the short-run effect because, in the short-run the firm responds only to a higher wage, but in the long-run, it responds both to a higher wage and to an increased capital stock that reduces marginal product of labor. Graphically, the additional effect in this example can be represented by a positive feedback loop.

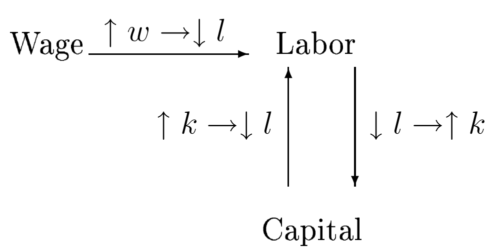

Figure 16.1: Positive Feedback Loop

Next, suppose that the two inputs, capital and labor, are *complements*.

Again, by the law of demand, if the wage increases, the firm will use less labor input, both in the short-run and in the long-run. Since the inputs are complements, the increased wage implies a reduced use of capital in the long run. Since $f_{kl} \geq 0$ for complements, the reduced capital used in the long-run will reduce the marginal product of labor, so in the long-run the firm will use still less labor. Again, we have a positive feedback loop.

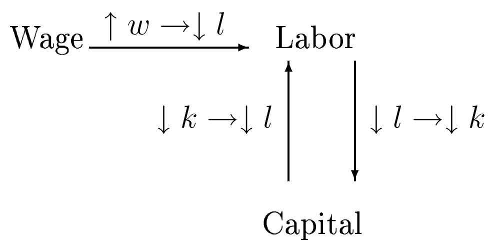

Figure 16.2: Positive Feedback Loop

The general positive feedbacks argument for two inputs (due to Milgrom and Roberts (1996)) goes as follows. Let X and Y be sublattices (for example, let $X = Y = \mathbb{R}$). Define:

$$x(y,t) = \arg\max_{x \in X} g(x,y,t)$$

and

$$y(t) = \arg\max_{y \in Y} g(x(y,t),y,t).$$

PROPOSITION 11: *Suppose that* $g : X \times Y \times \mathbb{R} \to \mathbb{R}$ *is supermodular, that* $t' \geq t$*, and that the maximizers described below are unique for the parameter values* t *and* t'*. Then:*

$$x(y(t'), t') \geq x(y(t), t') \geq x(y(t), t)$$

and

$$x(y(t'), t') \geq x(y(t'), t) \geq x(y(t), t)$$

Proof. By Topkis' Theorem, the function $y(t)$ is isotone ("weakly increasing"). Then, since $t' \geq t$, $y(t') \geq y(t)$. Similarly, by Topkis's Theorem, the function $x(y, t)$ is isotone. The claims in the theorems follow immediately from that and the inequalities $t' \geq t$ and $y(t') \geq y(t)$. □

Q.E.D.

Now let's apply the result. Let x be labor input, and y capital input, and let $t = -w_x$, where w_x is the price of labor. The firm's objective is to maximize:

$$g(x, y, t) = pf(x, y) - w_x x - w_y y.$$

If capital and labor are "complements" in the sense that $f_{xy} \geq 0$, then the firm's objective is supermodular in $(x, y, -w_x)$, because it verifies all the pairwise supermodularity conditions. Similarly, if capital and labor are "substitutes" in the sense that $f_{xy} \leq 0$, then the firm's objective is supermodular in $(x, -y, -w_x)$. We then have the LeChatelier Principle.

LeChatelier Principle *Suppose production is given by $f(k,l)$ where either $f_{kl}(k,l) \geq 0$ for all (k,l) or $f_{kl}(k,l) \leq 0$ for all (k,l). Then if the wage w increases (decreases), the firm's labor demand will decrease (increase), and the decrease (increase) will be larger in the long-run than in the short-run.*

Returning to our example above, consider the following three price vectors. The first two are p and p' as defined above and the third is $p'' = (2, 1.1, 1.1)$. In the example, the inputs act like substitutes when prices change from p to p' (the long-run input demand for good 3 rises with this increase in the price of good 2) but they act like complements when prices change from p' to p'' (the long-run input demand for good 2 falls with this increase in the price of good 3). It is this non-uniformity that enables the example to contradict the conclusion of the LeChatelier principle.

Chapter 17

Recovering the Production Set

We now turn to a set of questions concerning what can be learned from a set of observations of the form $\{(p_1, y_1), ..., (p_n, y_n)\}$. The questions are:

1. Is the set of observations consistent with profit maximization at fixed prices from some production set?
2. What can we infer about the underlying production set?
3. If the data set is sufficiently large, can we recover the entire production set? The production function?

For the first question, we can certainly infer from these observations that $\{y_1, ..., y_n\} \subset Y$. So, the firm's choices could be profit-maximizing only if for all n, m, $p_n \cdot (y_n - y_m) \geq 0$. If this inequality failed, then the choice made at prices p_n were less profitable than choosing the feasible alternative y_m.[1] Conversely, if all those inequalities are satisfied, then if the production set is $Y = \{y_1, ..., y_n\}$, then the choice is profit-maximizing for every price vector

[1] Notice that if this inequality fails, that does not establish that the firm is not a profit maximizer. An alternative explanation is that the firm is not a price taker.

p_n. These inequalities, then, characterize a dataset that is consistent with profit maximization.

For the second question, we can certainly infer that $Y^I = \{y_1, ..., y_n\} \subset Y$. We may call Y^I the *inner bound* on the production set. If we assume that the firm is maximizing profits, then the production set can only contain points for which the profits at prices p_n are no more than $p_n \cdot y_n$, that is, $Y \subset \{y | p_n \cdot y \leq p_n \cdot y_n\} = Y^O$, where Y^O is the *outer bound* on Y. Thus, $Y^I \subset Y \subset Y^O$.

If we assume that the production set satisfies free disposal, we can expand that inner bound. The free disposal inner bound is:

$$Y^I_{FD} = \{y | (\exists n) y \leq y_n\}.$$

The condition of free disposal implies that $Y^I_{FD} \subset Y$. With that extra assumption, the answer to the second question is $Y^I_{FD} \subset Y \subset Y^O$.

The third question, about large data sets is interpreted as supposing that we know the entire decision function $y(p)$. That implies that we know $\pi(p)$ because for each p, taking any $y \in y(p)$, $\pi(p) = p \cdot y$. By a previous proposition, we know that if Y is closed and convex and has the free disposal property, then $Y = \{y \in \mathbb{R}^n : p \cdot y \leq \pi(p) \text{ for all } p > 0\}$. So, in that case, the production set coincides with its outer bound $Y = Y^O$. Moreover, by the supporting hyperplane theorem, every point on the boundary of Y is chosen for some price vector, so in the limit, the inner bound also coincides with the production set: $Y = Y^I$. So, when Y is closed and convex and satisfies free disposal, the inner and outer bounds derived from $y(p)$ coincide with each other–$Y^I = Y^O$–and it follows that we can recover Y from the function $y(p)$ in that case.

Knowledge of the correspondence $y(p)$, however, is not sufficient when the production set Y when the set is not convex, even if we make the usual (and relatively innocuous) assumptions that Y is closed and satisfies free disposal. The outer bound, Y^O, being an intersection of closed half-spaces, is always convex, so it can't coincide with Y when Y is not convex. Hence, it cannot correspond to the inner bound, either. There is insufficient information in $y(p)$ to decide whether the points in the difference set $Y^O - Y^I$ are in the set Y.

When the production set is convex and production is described by a production function, the possibility of recovering the production set can take a particularly nice form.

Proposition 12: Duality of Profit and Production Functions *Suppose that $f(x)$ is a production function, that $\pi(p)$ is the associated profit function $\pi(p) = \max_x f(x) - p \cdot x$, where the price of output is normalized to one. If f is concave, then*

$$f(x) = \min_p \pi(p) + p \cdot x$$

Proof. By definition, for all x, p, $\pi(p) \geq f(x) - p \cdot x$, so $f(x) \leq \pi(p) + p \cdot x$ for all p. It follows that $f(x) \leq \min_p \pi(p) + p \cdot x$.

Assuming that f is concave, we now show the reverse inequality. Sine f is concave, then the production set $Y = \{(y, -x) | y \leq f(x)\}$ is convex. So, by the supporting hyperplane theorem, there exists some hyperplane that supports Y at the boundary point $(f(x), -x)$. That is, there is some $\widehat{p}$ such that for all x', $f(x') - \widehat{p} \cdot x' \leq f(x) - \widehat{p} \cdot x$. Then, $\pi(\widehat{p}) = f(x) - \widehat{p} \cdot x$, so $f(x) = \pi(\widehat{p}) + \widehat{p} \cdot x$. Hence, $f(x) \geq \min_p \pi(p) + p \cdot x$. $\square$

Q.E.D.

We have already seen that we cannot recover the production set or production function for the non-convex case, because there is necessarily a gap between the inner and outer estimates of the set. Hence, the preceding results give a complete answer to the question of when the production set can be recovered from sufficiently rich data about the choices of a competitive profit-maximizing firm.

Index