Lecture Notes on the Theory of Consumer and Producer Behavior

Marzio Galeotti

Manuali

Marzio Galeotti

Lecture Notes on the Theory of Consumer and Producer Behavior

Milano University Press

Lecture Notes on the Theory of Consumer and Producer Behavior / Marzio Galeotti. Milan: Milano University Press, 2026.

ISBN 979-12-5510-395-0 (print)

ISBN 979-12-5510-389-9 (PDF)

ISBN 979-12-5510-393-6 (EPUB)

DOI 10.54103/milanoup. 286

This volume, and Milano University Press publications in general, unless otherwise specified, are submitted to an external refereeing process under the responsibility of the Milano University Press Editorial Board. The published works are evaluated and approved by the Editorial Board, and must be compliant with the Peer review policy, the Open Access, Copyright and Licensing policy and the Publication Ethics and Complaint policy as reflected in MilanoUP publishing guidelines (Linee Guida per pubblicare su MilanoUP).

Published by Milano University Press
Via Festa del Perdono 7-20122 Milano
Web Site: https://milanoup.unimi.it
e-mail: redazione.milanoup@unimi.it
The print edition of this volume can be ordered from bookstores, and is distributed by Ledizioni (http://www.ledizioni.it)

General note on figures and tables
Unless otherwise indicated, figures and tables have been prepared by the author(s) of the chapter in question. The Editor is committed to striving to satisfy all copyright requirements concerning graphics, images and tables for which it has not been possible to identify the type of licence used.

Table of Contents

Foreword

These notes represent the material that I have been using over the years to teach first-year PhD students in Economics at the University of Bergamo, of Pavia, and of Milan, as well as Master in Economics students at Bocconi University. Though they have significantly evolved over the years alongside my research, their core ultimately originates from the very first notes I took as a PhD student myself when attending Microeconomics 1 at New York University, now many years ago. The content of that course, and in particular on producer behavior, became the background of my dissertation work on the theory of investment and dynamic factor demands. One specific production factor was energy, and the knowledge I had accumulated on that topic became instrumental for the interest and passion and research activity that I subsequently developed in Environmental Economics, and in particular in the Economics of Climate Change.

When I arrived at NYU from Bocconi University, with a degree in Economics and Political Sciences and a thesis in International Finance focused on capital movements and the balance of payments, I could not have foreseen that I would have radically changed the main direction of my research. The NYU courses in Micro and Macroeconomics were instrumental to changing the trajectory of my interests. Undoubtedly, this has also been thanks to the mentorship of my teacher and advisor Mark A. Schankerman, now at LSE, and of the leadership of M. Ishaq "Ned" Nadiri, together with his group, which included among others Ingmar Prucha, Jeff Bernstein, Pierre Mohnen, and my longtime friend Angelo M. Cardani.

Finally, I am grateful to the Department of Environmental Science and Policy at the University of Milan and to Fondazione Eni Enrico Mattei for providing an ideal environment where to carry out the last part of this effort.

Chapter 1

Introduction

These notes deal with the economic behavior of the individual consumer and of the individual producer. To represent the foundational role of this theory one can refer to perhaps the first topic of a standard Introductory Economics course, which is the notion of Economic System. It is typically illustrated with the help of the so-called Circular Economic Flow diagram reproduced in Figure 1.1.

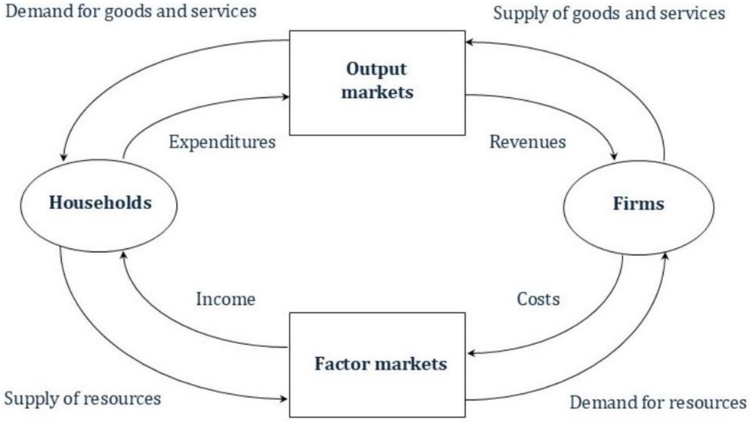

Figure 1.1 *Circular economic flow.*

In the most elementary representation of how a market economy works there are two groups of agents, households and firms. Households perform two basic economic functions: consume goods and services and make use of the resources – typically services – they own. Goods and services are bought – and therefore demanded – in appropriate markets for goods and services; factor services are sold –

and therefore supplied – in appropriate markets for production factors. Firms carry out the fundamental function of transforming resources into other resources. Better stated, firms transform factor services into manufactured goods and services. To that end they acquire – and therefore demand – services in factor markets. These services enter as inputs the production process which yields outputs of goods and services sold – and therefore supplied – in goods markets.

Leaving aside other aspects, however relevant, the circular flow scheme describes the aggregate behavior of the two fundamental actors – the consumer and the producer – and highlights two crucial aspects of the economic system: agents' decisions over goods and services and market exchange.

A second illustration, logically connected to the one above, often discussed at the outset of a Microeconomics course is the plan of work, shown in Figure 1.2.

Moving from bottom to top, in essence the economy can be viewed as the set of markets where all goods and services – inputs and outputs – are traded and where price and quantity of each one of them are established. In each individual market the price and quantity of a good or service results from the equilibrium between demand for it and supply of it. Market demand and supply for each good and service in turn stem from the aggregate choices of consumer-demanders and producers-suppliers. At the root of this process is therefore the study of individual decision making, the determinants of the behavior of the representative consumer and of the representative producer.

These notes provide a fairly advanced illustration of the neoclassical theory of the behavior of the individual consumer and of the individual producer.

***

In terms of the sequence displayed in Figure 1.2, in these notes it is discussed only what is inside the two upper boxes on the left and on the right of the first panel. Consumer and producer theory are the bedrock foundation on which so many theoretical structures in Economics are built. While the list cannot be exhaustive, modern applied Macroeconomic analysis (based on Dynamic Stochastic General Equilibrium models and on Computable General Equilibrium models), Economics of growth, Labor economics (labor demand and supply), Finance (theory of saving), Energy economics (energy demand and supply), Environmental economics (externalities, public goods, environmental policies), International trade are areas with foundations in the present theory of consumer and producer behavior.

***

While much of the discussion on the methodological aspects underpinning the theory expounded here is taken for granted, a few assumptions concerning the conceptual framework are worth recalling explicitly. Thus:

- Agents are rational and self-interested;
- Agents optimize, i.e. they pursue the highest return on their choices;

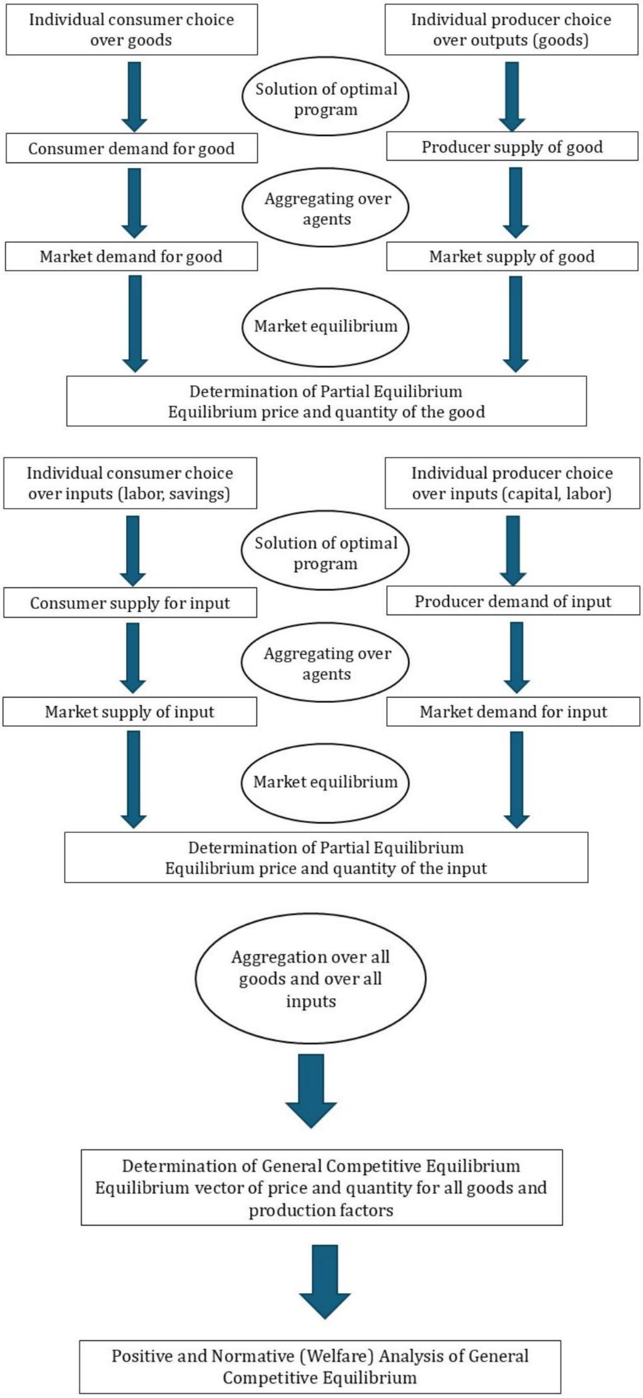

Figure 1.2 *Microeconomics sequence.*

- Agents are constrained in their decisions, as resources – including time – are scarce or finite;
- Agents are fully informed, information is free, uniform, and complete;
- The world is one of certainty: there is nothing between a decision and a choice, they are the same thing.

In addition:

- The analysis is a-temporal (static): on the consumption side there is no distinction between durable and non-durable consumption goods (goods considered here are all non-durable); on the production side there is no capital investment and there is no technological change;
- Agents are price-takers: no consumer or producer has market power to any degree;
- Agents interact with each other only through prices: an agent's consumption or production does not affect directly the welfare of other agents. No externalities;
- For all markets in the economy: producers maximize profit, given prices and technology; consumers maximize utility, given prices and preferences; markets clear (equilibrium) so that demands and supplies are equal and equilibrium prices and allocations are determined.

***

For both consumer and producer, the exposition of the theory logically proceeds as follows. First, we discuss the constraint(s) that limit the agent's freedom of choice. Second, we lay out the objective/goal the agent aims to pursue. Third, we describe the optimal choices of the agent. Mathematically, these three steps amount to solve a constrained optimization problem whose constituent elements are the agent's objective function and any constraint she is subject to. The solution yields the optimal decisions rules whereby the chosen (endogenous) variables representing the agent's decision are expressed in terms of outside (exogenous) variables, which are given to the agent. This represents the agent's equilibrium.

The decision rules serve the purpose of explaining and predicting the behavior of the agent. The second part of the theory is about predicting that behavior when one or more exogenous variables ("the economic environment") change. The agent will re-optimize reaching a new equilibrium: the analysis of the change, comparing the "initial" and "final" equilibria, is referred to as comparative statics.

The third and final part of the discussion deals with the characterization of the agent's optimized objective, which describes the best outcome the agent can obtain when her behavior is optimal, as described in the previous parts of the analysis. It turns out that also the agent's optimized objective is affected by the above-mentioned exogenous variables, so that it is of interest to ask how the optimized objective responds to changes in the exogenous variables.

***

The above-described approach is followed in the first two chapters of this volume. Technically speaking, the agent's optimal behavior is represented as a "primal" problem, such as a consumer maximizing utility given a budget constraint, or a firm maximizing profit given technology. Here, both graphical and mathematical tools will be used in the exposition of the theory in order to help the reader connect with the former approach of introductory/intermediate courses and the latter approach proper of advanced treatments. The remaining two chapters will again deal with consumer and producer behavior but from a "dual" perspective. This is the rephrased problem, such as a consumer minimizing the expenditure to achieve a certain level of utility, or a firm minimizing cost to produce a specific output. Here, only the mathematical method will be used. The duality approach has several theoretical applications, e.g. in index number theory and cost-of-living indexes, choices under rationing and calculation of shadow prices which represent the economic value of constraints or resources, welfare economics and notions of consumer surplus (compensating and equivalent variation), but it has proved mostly useful in the empirical application of consumer and producer theory.

Chapter 2

Consumer behavior: preferences, utility and choice

2.1 Consumption set

Let $q = (q_1, q_2, \ldots, q_n)$ be a consumption bundle or basket consisting of a combination of quantities of n goods (or services) $q_i \geq 0$ $(i = 1, \ldots, n)$. Quantities are taken to be non-negative. We use subscripts to denote individual goods and superscripts to denote individual bundles.

Consider a set C of (physically) feasible consumption bundles q. An example is given in Figure 2.1.

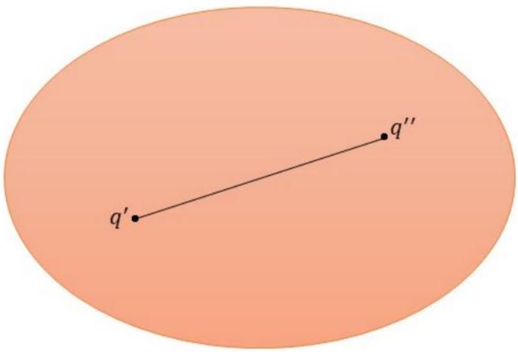

Figure 2.1 *Consumption set.*

In the figure q' and q'' represent two consumption bundles. Note that this set excludes a number of bundles that the agent cannot actually consume. An example is a bundle one element of which is the number of trips to Mars. As of today, such a bundle is not feasible, i.e. it is not physically consumable.

We assume that the consumption set C has the following properties:

C.1 C is **bounded** from below: $C = \{q\}$ where $q = (q_1, q_2, ..., q_n) \geq 0$;

C.2 The components of C are **infinitely divisible**;

C.3 **Additivity**: if q' and $q'' \in C$, then $(q' + q'') \in C$;

C.4 C is **closed** with respect to multiplication: if $q \in C$ then $\lambda q \in C$, $\forall \lambda \geq 0$.

Owing to C.2 we will assume throughout that bundles are characterized by quantities that are continuous variables. This assumption is made for convenience to enable the use of standard mathematical techniques. In reality there are several examples of discrete goods (e.g. shoes). A theory of consumer choices when goods are discrete can be constructed, see Varian (2014).

Properties C.1-C.4 imply C.5:

C.5 **Convexity of** C: if q' and $q'' \in C$, then the weighted average $\overline{q} = \lambda q' + (1-\lambda) q'' \in C$ for $\forall 0 \leq \lambda \leq 1$.

Proof. Consider $q', q'' \in C$. From C.4 we have that $\lambda q' \in C$ and $(1-\lambda) q'' \in C$. If $0 \leq \lambda \leq 1$ then from C.3 it follows that $\lambda q' + (1-\lambda) q'' \in C$. Otherwise, it could be that $q'' < 0$ which is not possible because of C.1.

Math digression: Convex sets

A set is convex if, given any two points in the set, the set contains the whole line segment that joins them. In Figure 2.2 we represent a convex (left) and a non-convex (right) set in two-dimensional space.

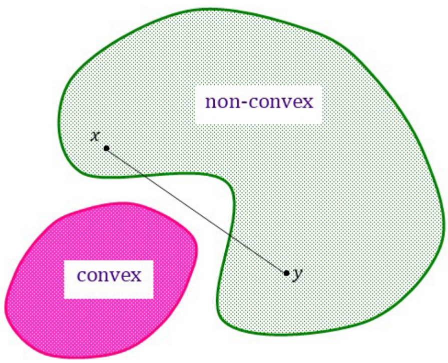

Figure 2.2 *Convex and non-convex sets.*

It can be seen from the figure on the right that if we draw the cord between the two extreme points x and y there are average points that do not belong to the set. Figure 2.3 shows examples of convex and non-convex sets.

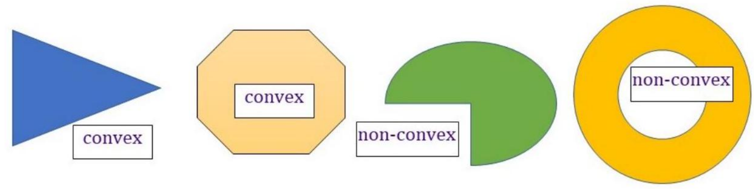

Figure 2.3 *Examples of convex and non-convex sets.*

We will often make use of the 2-goods case. The consumption set is shown in Figure 2.4.

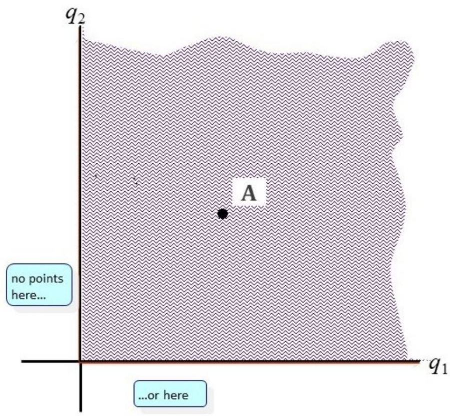

Figure 2.4 *Consumption set in 2-good case.*

Every point in the positive orthant, like A, is a consumption bundle. Note that the quantities of some good within a bundle can be zero, so that also the points on the axes belong to the consumption set. It can be verified that the set is convex: see Figure 2.5.

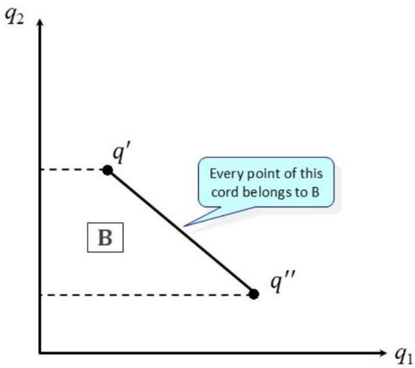

Figure 2.5 *Convex consumption set in 2-good case.*

Summarizing, **a physically feasible consumption set is bounded from below, contains zero and is convex.**

2.2 Budget set and budget constraint

The consumption set describes all bundles the consumer can actually consume. We now partition the set into the subsets of bundles the consumer can afford to buy and thus consume and of bundles she cannot afford. The former is referred to as the **budget set**.

We make three assumptions to restrict the feasible consumption set to an economic meaningful one:

B.1 The agent has a fixed income I;

B.2 The agent faces fixed commodity prices p;

B.3 The agent has a budget constraint given by the inequality $\sum_{i=1}^{n} p_i q_i \leq I$.

Note that we make here two assumptions. The first one is that monetary income I is exogenously given to the agent. This implies that we assume that her decisions do not include how to generate or allocate income. If we were to extend the framework to incorporate labor supply decisions, then at least a portion of income – wage income – would be endogenously determined by the agent herself together with consumption decisions. Similarly, capital income resulting from saving decisions would be endogenous. We assume away these possibilities to keep the analysis simple and focused on consumption decisions. The second assumption is that the consumer is price-taker, according to B.2.

In the space with (just) two goods, the restricted feasible set is represented as follows.

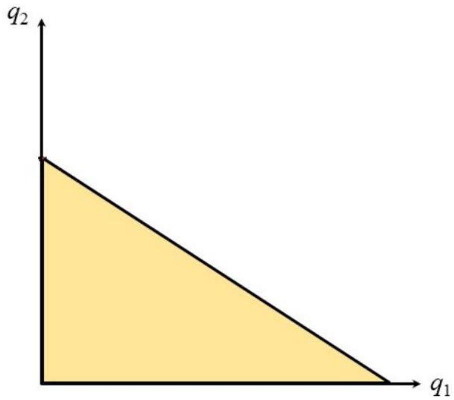

Figure 2.6 *Restricted feasible consumption set.*

This set is known as the budget set. It includes all feasible consumption bundles that the consumer can afford (to buy), given her income and the prices of the two goods. In the interior of the set some income is not spent, whereas bundles outside the budget set correspond to expenditures that exceed income and that the agent cannot therefore afford. The frontier of the set is referred to as the **budget line** or **budget constraint,** and it represents the set of bundles for which the consumer spends all her income. In the two-good space, it is given by $p_1 q_1 + p_2 q_2 = I$. More generally $\sum_{i=1}^{n} p_i q_i = I$.

If prices were not fixed, or if the consumer was able to affect prices with her behavior, then the budget constraint would no longer be linear and it would be written as $I = p_1(q_1)q_1 + p_2(q_2)q_2$, so that:

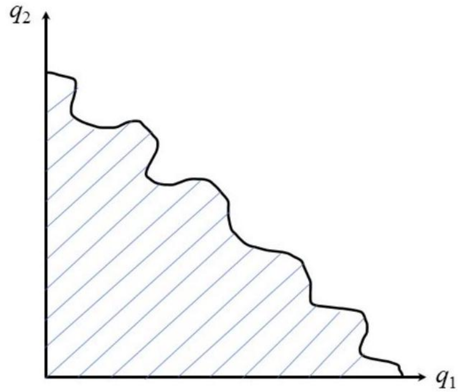

Figure 2.7 *Non-linear budget set.*

One important assumption underlying the budget set and the budget constraint is that the agent cannot spend more than her income. Therefore, in this framework she cannot borrow. In fact, borrowing would make necessary a dynamic/multiperiod analysis as what she borrows today she must give back tomorrow. Thus, if she consumes more than her income today, she will have to do the opposite tomorrow.

An additional implication is that in this framework it makes no sense (it is not rational) not to spend all income today, as saving (consumption lower than income) is fruitless, is like money thrown away. It follows that **the optimal decision by the consumer will be a bundle of goods that always lies on the budget line**.

The **slope of the budget constraint** can be seen by solving the budget constraint explicitly for the quantity of one good: $q_2 = \frac{I}{p_2} - \frac{p_1}{p_2} q_1$. Thus, the slope is $-\frac{p_1}{p_2}$ and the intercepts with the axes are $\frac{I}{p_2}$ and $\frac{I}{p_1}$. These intercepts represent the amount of one good that could be purchased if all income were spent in that good. See Figure 2.8.

It is important to understand how the budget line changes when income and/or one or all prices change. There are several cases.

1. One price, say p_1, increases/decreases;
2. There is an increase/decrease in money income I;
3. All prices increase/decrease by the same proportion θ;
4. All prices and income increase/decrease by the same proportion θ.

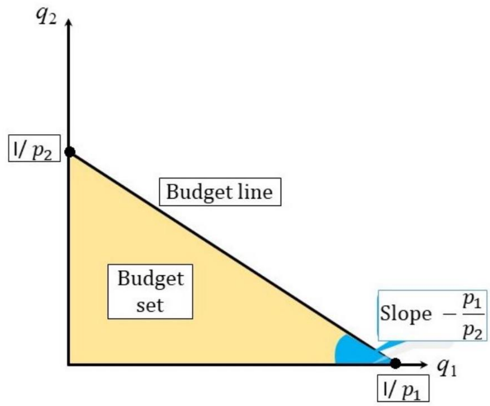

Figure 2.8 *Budget line.*

When the price of good 1 p_1 increases (resp. decreases) the budget line becomes steeper (resp. less steep) and rotates around the intercept with the vertical axis (as I/p_2 is unaffected). Similarly for changes in p_2. An increase (resp. decrease) in money income I shifts the budget line outward (resp. inward) with unchanged slope, thus it moves in a parallel way away from (resp. toward) the origin. It should be clear that in both cases the budget set expands (resp. shrinks) so that more (resp. fewer) bundles are now affordable. These two cases are shown in the figure below.

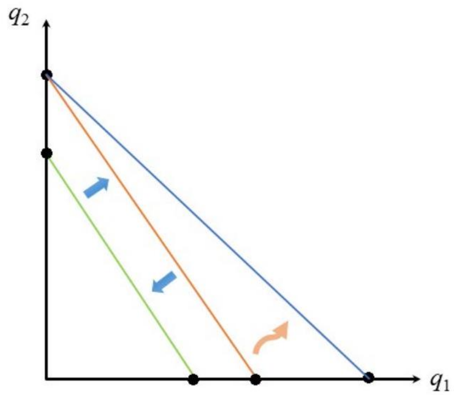

Figure 2.9 *Changes in the budget line.*

What happens if all prices increase (resp. decrease) by the proportion θ? The new budget constraint is $\theta p_1 q_1 + \theta p_2 q_2 = I$ for $\theta > 0$. It is easy to see that this is equivalent to a reduction (resp. increase) in income: $p_1 q_1 + p_2 q_2 = I/\theta$. Finally, what happens if all prices and income increase (resp. decrease) by the same proportion θ? Nothing: indeed under full indexation of all monetary variables $\theta p_1 q_1 + \theta p_2 q_2 = \theta I \implies p_1 q_1 + p_2 q_2 = I$ and the position of the budget line in terms of slope and intercepts remains unchanged.

2.3 Preference ordering and the utility function

What is a **preference ordering**? It is the ability of the consumer to rank all the feasible bundles of goods she faces according to some predetermined set of criteria from which implications for her behavior can be drawn.

Let $\succ$ and $\sim$ be binary relations defined over the set C where $q' \succ q''$ means that the agent strictly prefers bundle q' to q'' and $q' \sim q''$ means that she is indifferent among the two (with $\succcurlyeq$ if she weakly prefers).

To construct a theory of consumer choice we need to impose some structure on the agent's preferences. This takes place through so-called **axioms** on preferences, i.e. fundamental statements that are not subject to confutation. Thus, they must be general (universal) and reasonable (acceptable).

There are six axioms, the first four are fundamental.

A.1 **Completeness**. For any bundle q' and $q'' \in C$, either $q' \succ q''$ or $q' \prec q''$ or $q' \sim q''$. This means that the consumer has always preferences over bundles and she is able to always express a ranking between any pair of them.

A.2 **Reflexivity**. For every bundle q' we have that $q' \succ q'$. This is a technical/mathematical requirement.

A.3 **Transitivity**. For bundles q', q'' and $q''' \in C$, if $q' \succ q''$ and $q'' \succ q'''$ then it must be that $q' \succ q'''$. This means that one's preference ranking does not depend on the order of the bundles.

If axioms A.1-A.3 hold, then a preference ordering is defined, i.e. the consumer can order all bundles in a defined rank.

Working with preference ordering does not get us far. We need to be able to use familiar mathematical tools, such as algebra and calculus. We must then be able to translate a preference ordering onto the space of real numbers.

In order to be sure that there is a mapping between preference ordering and the real line we need to make sure that there are no discontinuities. We add the following axiom.

A.4 **Continuity**. Let $\left\{q \in \overline{C} | q^0 \succ q\right\}$ be any feasible consumption set where $q^0 \succ q$. Then, for any $q^0 \in \overline{C}$, the sets $\left\{q \in \overline{C} | q^0 \succ q\right\}$ and $\left\{q \in \overline{C} | q^0 \prec q\right\}$ are both closed sets. These are the "preferred" and the "non-preferred" sets. Recall that a closed set is a set that contains its own boundaries: if $q' \succ q''$ and we change infinitesimally q' this does not imply that $q'' \succ q'$. Thus, in the neighborhood of q' there are no jumps. Graphically:

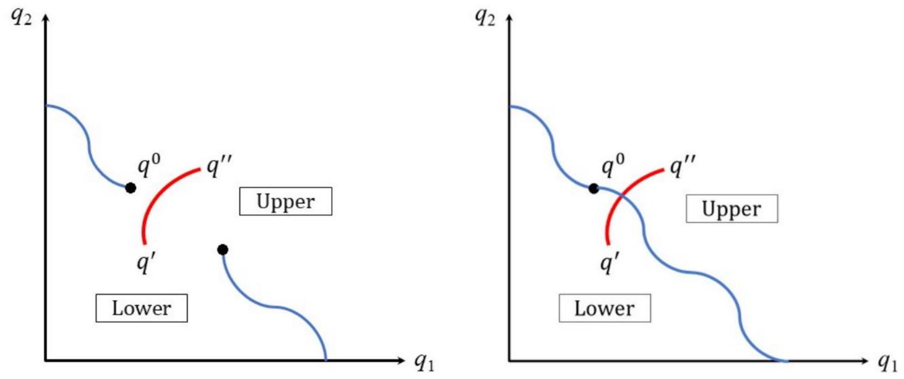

Figure 2.10 *Continuous and non-continuous preferences.*

Axioms (A.1)-(A.4) allow proving a fundamental **theorem on the existence of a utility function** $U(q)$ defined over bundles of goods in C such that there is a mapping from the space of preferences onto the real line. That is:

$$q' \succsim q'' \Leftrightarrow U(q') \geq U(q'')$$

$$q' \precsim q'' \Leftrightarrow U(q') \leq U(q'')$$

That is, a preference for a bundle over another one corresponds to a real number associated with the first bundle higher than the real number associated with the second bundle.

Proof. See Varian (1979).

The Continuity axiom A.4 implies **continuity of the utility function**.

Graphically, in the three-dimensional space the utility function is the outer frontier of a surface that looks like the following:

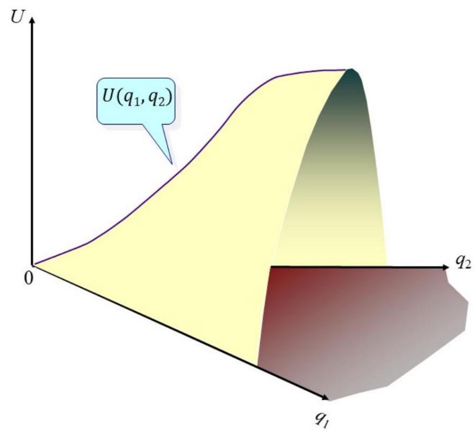

Figure 2.11 *Utility function.*

Suppose we fix q_2. Then we can also represent the utility funtion in the bidimensional space as follows:

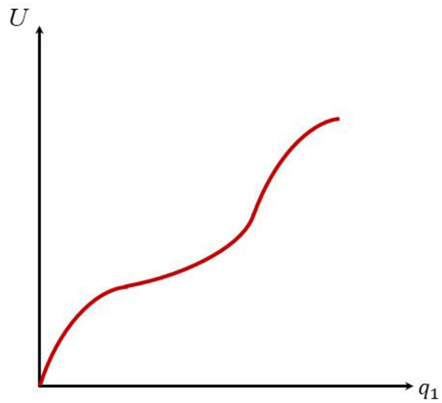

Figure 2.12 *Utility function.*

Note that the utility function has the following properties:

U.1 U is **monotonic**. This means that $q' \gtrsim q'' \Rightarrow U(q') \geq U(q'')$. To a preferred bundle is associated a higher real number.

U.2 U is **not unique**. More precisely, it is unique up to any monotonic transformation.

Proof. Define a monotonically increasing function F, i.e. such that $F' > 0$. Then, we have that $U(q') \geq U(q'') \Rightarrow F(U(q')) \geq F(U(q''))$. Now call $F(U(q)) = \Phi(q)$, then $\Phi(q') \geq \Phi(q'')$ which is a valid utility function.

The non-uniqueness of the utility function implies **ordinality** as opposed to cardinality: the values that U takes on have no meaning and need not have it, as the only thing that matters is that the value of utility that is strictly higher than another one means that the former bundle of goods is strictly preferred to the latter.

Example. A well-known type of preference ordering are **lexicographic preferences**. They are defined by the following. Consider $q' = (q'_1, q'_2, ..., q'_n)$ and $q'' = (q''_1, q''_2, ..., q''_n)$. Then:

a. If $q'_1 > q''_1$ then $q' \succ q''$

b. If $q'_1 = q''_1$ then $q' \succ q''$ if $q'_2 > q''_2$

c. If $q'_1 = q''_1$ and $q'_2 = q''_2$ then $q' \succ q''$ if $q'_3 > q''_3$

and so on.

While defining a preference ordering (axioms A.1-A.3 hold), lexicographic preferences do not define a continuous set and therefore cannot define a utility function.

The next axiom is:

A.5 **Global non-satiation**. For every $q^0 = (q_1^0, q_2^0, ..., q_n^0)$ there exists some q' such that $q' \succ q^0$ where $q' = (q_1^0 + \epsilon, q_2^0, ..., q_n^0)$ with $\epsilon > 0$. Thus, **utility is strictly increasing in the quantity of any one good**: this leads to the positive marginal utility of anyone good.

Notes. (1) In many cases, the theory presented in these notes involves weak inequalities. Whenever that is the case, however, for convenience we will assume inequalities to be strict (e.g. utility does not decrease if the quantity of one good increases → utility increases if the quantity of one good increases). (2) Satiation can be local. In such case we cannot rule out satiation or saturation points where utility decreases when the quantity of a good (e.g. an ice cream) further increases. We will assume such bundles away (see Figure 2.18).

After having established continuity and monotonicity properties for the utility function, we turn to curvature.

A.6 **Convexity**. Let $q' \succ q''$. Define an "average" bundle $\overline{q} = \lambda q' + (1-\lambda)q''$. Then $\overline{q} \succ q''$ for every $0 \leq \lambda \leq 1$.

Note. This is strict convexity. Weak convexity implies replacing the weak preference ranking. In particular, if we have $q' \sim q''$ then $\overline{q} \gtrsim q''$.

The convexity axiom implies that the **utility function is quasi-concave**.

Math digression: Curvature and Quasi-concavity

A (univariate) function $f(x)$ is said to be **quasi-concave** if, for any two values x_0 and x_1 such that $f(x_0) \geq f(x_1)$ and for $0 \leq \lambda \leq 1$, it must be that $f(\lambda x_0 + (1-\lambda)x_1) \geq f(x_0)$. Graphically, we have:

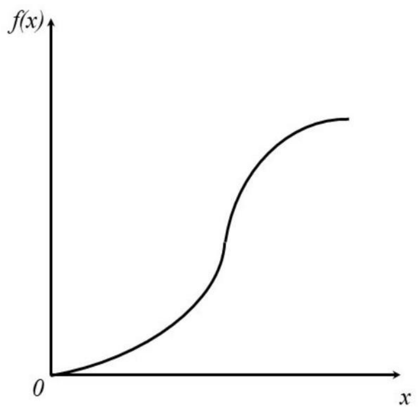

Figure 2.13 *Quasi-concave function.*

Note: (1) Graphically, quasi-concavity implies that, given a reference value, the function cannot decrease (can have a horizontal, concave, convex, or S-shaped portion) relative to that value; (2) The above definition is that of weak quasi-concavity; (3) if $f(\lambda x_0+(1-\lambda)x_1) \le f(x_1)$ for $f(x_0) \ge f(x_1)$, **the function is said to be (weakly) quasi-convex**.

A stronger form of curvature is a function $f(x)$ being **concave** if, for any two values x_0 and x_1 such that $f(x_0) \ge f(x_1)$ and for $0 \le \lambda \le 1$, it must be that $f(\lambda x_0 + (1-\lambda)x_1) \ge \lambda f(x_0) + (1-\lambda) f(x_1)$. Graphically, we have:

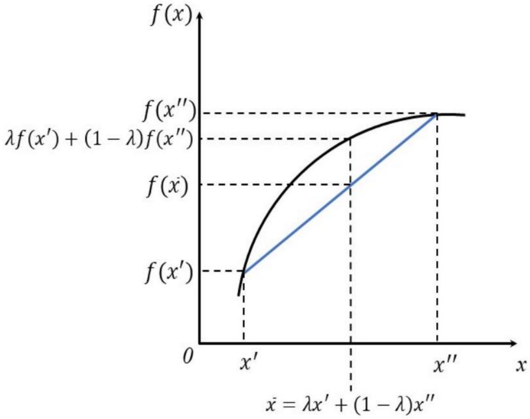

Figure 2.14 *Concave function.*

<u>Notes</u>. (1) Graphically, concavity implies that, given a reference value, the function has always to increase relative to that value; (2) The above definition is that of weak concavity; (3) If $f(x)$ is concave, then the function $-f(x)$ is said to be **(weakly) convex**; (4) a concave function is also quasi-concave, but not vice versa.

In the case of multivariate functions $f(x_1, x_2, \ldots, x_n)$ the concavity/convexity property has to be checked by looking at the principal minors of the matrix of second partials, called Hessian matrix H :

$$H = \begin{bmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{1n} & \cdots & f_{nn} \end{bmatrix}$$

For strict concavity (weak concavity) we must have: $|H_1| < 0; |H_2| > 0, |H_3| < 0, \ldots, |H_n| \gtrless 0$, alternating signs depending on whether n is odd: negative (non-positive) or even: positive (non-negative). Thus, they alternate in signs.

For strict convexity (weak convexity) instead we must have: $|\mathrm{H}_1| > 0, |\mathrm{H}_2| > 0, |\mathrm{H}_3| > 0, \ldots\ldots\ldots, |\mathrm{Hn}| > 0$ always positive (non-negative).

The conditions for quasi-concavity/convexity are quite different. We have to take the Hessian matrix and add a border with the first partials of the function, called Bordered Hessian matrix B :

$$B(f,g) = \begin{bmatrix} 0 & \frac{\partial g}{\partial x_1} & \frac{\partial g}{\partial x_2} & \cdots & \frac{\partial g}{\partial x_n} \\ \frac{\partial g}{\partial x_1} & \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial g}{\partial x_2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g}{\partial x_n} & \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

For strict quasi-concavity (weak quasi-concavity) we must have: $|B_2| < 0; |B_3| > 0, \ldots, |B_n| \lesseqgtr 0$, alternating signs depending on whether n is even: negative (non-positive) or odd: positive (non-negative). B2 is the first 2×2 principal minor. Thus, they alternate in signs but note that is the opposite relative to concavity.

For strict quasi-convexity (weak quasi-convexity) instead we must have: $|B_2| < 0; |B_3| < 0, \ldots, |B_n| < 0$ always negative (non-positive). Note that this case is just the opposite relative to convexity.

The following important concept relates to the level curves of a utility surface and its curvature.

2.4 Representation of preferences and utility

An **indifference curve** (IC) is the boundary of the set $\{q | U(q) = k\}$ for some arbitrary k. In other words, it represents the set of bundles which are indifferent to one another.

Graphically, indifference curves correspond to the level curves of the utility surface. In the two-good space:

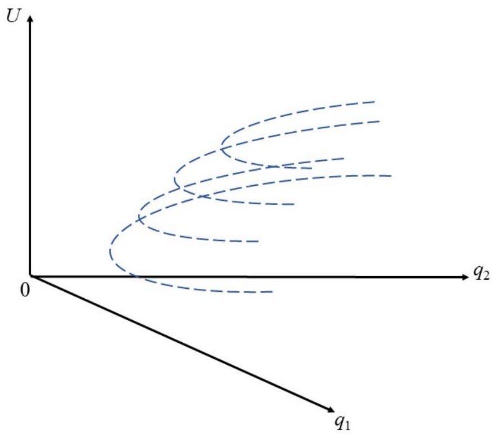

Figure 2.15 *Indifference curves.*

ICs are characterized by a few properties which are implied by the axioms on preferences. Specifically:

I.1 **ICs are downward sloped**. Consider two bundles $(q^0, q') \in \overline{C}$ such that $\{q^0|U(q^0)=k\}$ and $\{q'|U(q')=k\}$. Then, given q^0 we have both $\left\{q' \notin \overline{C}|q^0 \succ q'\right\}$ and $\left\{q' \notin \overline{C}|q^0 \prec q'\right\}$.

Proof. Consider the two-good space and divide it into four regions, as shown in Figure 2.16. Given a bundle q^0, by axiom A.4 (continuity) there exists a bundle q' which is indifferent to q^0. That is, $q^0 \sim q'$. By axiom A.5 (non-satiation) q' cannot lie in regions A or C, the preferred and non-preferred sets. Otherwise, we would have to have that $q' \succ q^0$ (region C) or $q' \prec q^0$ (region A). Hence bundle q' must be either in region B or region D, which are the set of indifferent bundles. Thus, an indifference curve slopes downward.

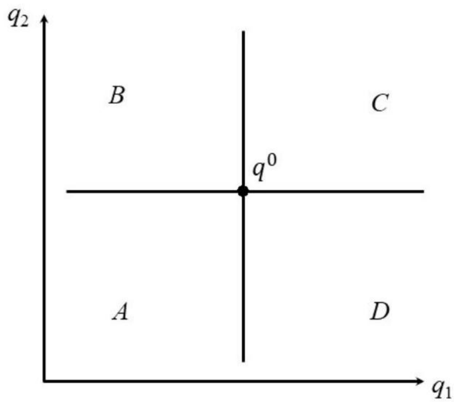

Figure 2.16 *Indifference curves slope downward.*

I.2 **ICs do not intersect each other.** For utility levels k and k' define the sets $\{q|U(q) = k\}$ and $\{q'|U(q') = k'\}$. Then, the intersection of q and q' is the empty set ($q \cap q' \equiv \varnothing$).

Proof. By definition of indifference curve, in Figure 2.17 we have that bundles $B \sim A$ and $A \sim C$. But, by axiom A.3 (transitivity), it must be that $B \sim C$ and, at the same time, by axiom A.5 (non-satiation) $B \succ C$, which leads to a contradiction. Hence, ICs cannot intersect each other.

I.3 **ICs are associated to higher utility levels the more distant from the origin they are**. The set of indifferent bundles q $\{q|U(q) = k\}$ that are strictly preferred to the set of indifferent bundles q' $\{q'|U(q') = k'\}$ are such that $k > k'$.

Proof. See Figure 2.15.

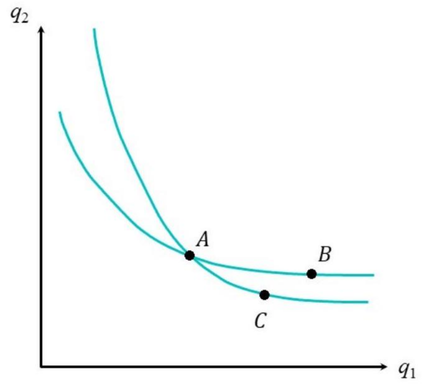

Figure 2.17 *Indifference curves do not intersect each other.*

As noted before, we do not need a cardinal utility but just an ordering of preferences. The important thing is that preferred bundles lie on an indifference curve that is higher than the one of less preferred bundles. The value of the associated utilities does not matter.

Property I.3 is implied by axiom A.5 about global non-satiation. This excludes the possibility of satiation points, where additional consumption of one good – say, ice cream – leads to a reduction, rather than a further increase, in utility. This possibility is shown in Figure 2.18.

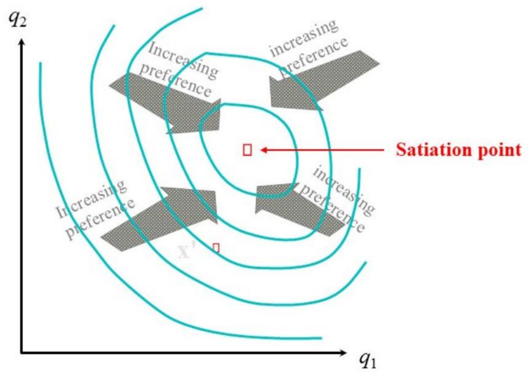

Figure 2.18 *Satiation points.*

If there are such bundles we can assume that the agent will make her "typical" consumption decisions well away from those points, in which case we would replace A.5 with a "local non-satiation" axiom.

We can consider two additional preference possibilities.

One commodity whose consumption does not generate utility to the agent, but rather dis-utility, is called a Bad. A bad is a commodity the agent does not like. Traditionally labor is considered a bad and leisure a good. More significantly, pollution is a bad. Assuming that there are two items in a bundle, one good (carrots in a soup) and one bad (beans in the soup), if we give the consumer more of the bad, to keep her on the same indifference curve we have to give her some extra good to compensate her for having to put up with the bad. Indifference curves in that case slope up and to the right as depicted in the left panel of Figure 2.19.

A good is a Neutral good if the consumer doesn't care about it one way or the other. If a consumer is just neutral about beans in the soup her indifference curves will be vertical lines as depicted in the right panel of Figure 2.19. She only cares about the quantity of carrots she has in her soup and does not care at all about how many beans she has. The more carrots the better but adding more beans does not affect her at all.

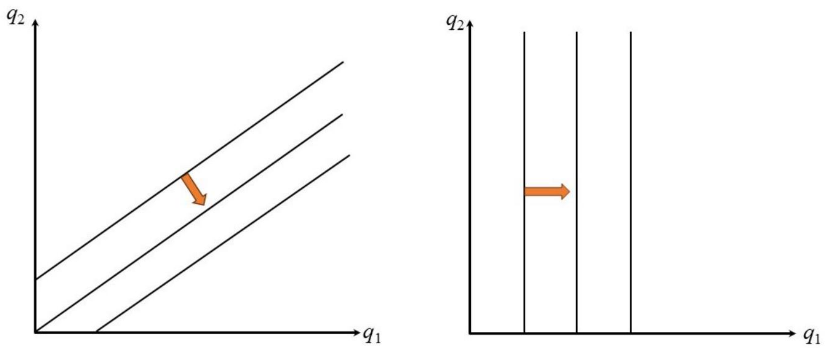

Figure 2.19 *Bads and Neutral goods.*

I.4 **ICs are (weakly) convex with respect to the origin**.

Convexity of ICs is implied by the quasi-concavity of the utility function. More precisely, upper contour sets of the utility surface are convex. Note that convexity of indifference curves includes linear ICs or L-shaped ICs, but not concave ICs with respect to the origin or ICs which include portions that are concave. See Figure 2.20.

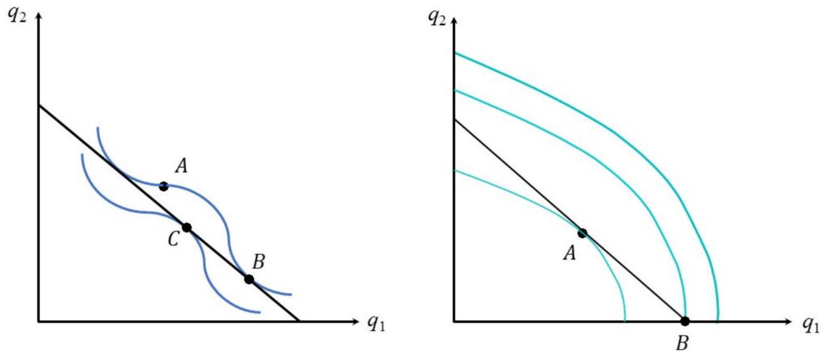

Figure 2.20 *Non-convex indifference curves.*

2.5 Marginal utility and Marginal rate of substitution

Individual preferences can be characterized both graphically – as done above – and analytically. Consider the following question: if the consumer wants to change the quantity consumed of every good, what is the resulting change in utility? The answer is given by the total differential. Consider the utility function $U = U(q)$ and take the total differential represented as follows:

$$dU = \sum_{i=1}^{n} \omega_i dq_i \tag{1}$$

Thus a marginal change in utility is brought about by a weighted average of marginal changes in the quantities of the various goods. We will consider two special cases of (1).

<u>*Special case 1*</u>. Let the quantity of only one good change. Thus: $dq_j = 0 \;\; \forall j \neq i$. From (1) we get:

$$\omega_i = \frac{\partial U(q)}{\partial q_i} = MU_i(q) \tag{2}$$

Each "weight" ω_i represents the change in utility brought about by a change in the quantiy of a good, holding all remaining quantities constant. This notion is referred to as the **Marginal utility of a good**. It is easy to see that the marginal utility is positive (non-negative) owing to axiom A.5 (non-satiation) or, equivalently, to the quasi-concavity of the utility function. We can therefore rewrite (1) as follows:

$$dU = \sum_{i=1}^{n} MU_i(q) dq_i \tag{3}$$

Note that marginal utilities are not constant but functions of q. This is represented in the two-good case in Figure 2.21.

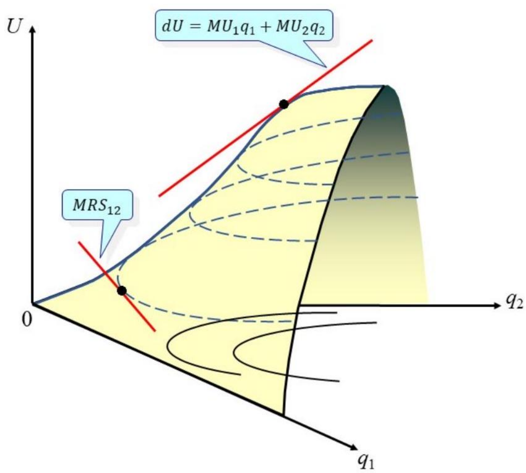

Figure 2.21 *Change in utility.*

Special case 2. We now let only two goods change, holding the quantities of all the other goods unchanged. In addition, we force utility to remain unchanged. Therefore: $dU = dq_k = 0 \;\; \forall k \neq i, j$. Thus, (1) becomes:

$$0 = U_i dq_i + U_j dq_j \rightarrow -\left(dq_j / dq_i\right) = MU_i(q)/MU_j(q) = MRS_{ij}(q) \tag{4}$$

The **Marginal rate of substitution** (MRS) between pairs of goods represents the rate at which the consumer is willing to exchange one good for the other. Because utility cannot change, one good can only be substituted for the other. It is equal to the ratio of the marginal utilities. Note that also the MRS is not constant but a function of q. Graphically, the MRS is the tangent at a given indifference curve at a point.

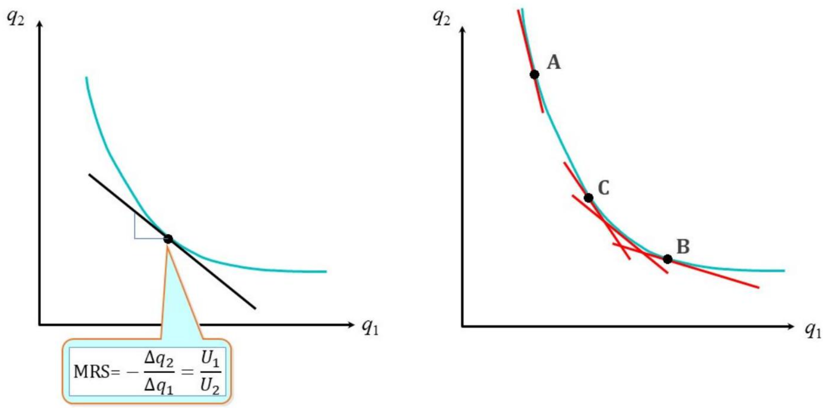

Figure 2.22 *Marginal rate of substitution.*

Note that the strict convexity of ICs implies that the MRS is decreasing, as the consumer becomes more reluctant to give up a quantity of the good which is less abundant, so that she pretends a larger quantity of the other good.

The property of ICs being negatively sloped implies substitution between the two goods in the two-good case. In addition, because ICs are weakly convex, they can contain linear portions or be uniformly linear in which case the MRS is a constant value: it does not depend on the relative quantity of the goods and it has same value at each point of the given indifference curve.

2.6 Properties of the utility function

We can now state all properties the utility function possesses.

U.1 U is a **real-valued and continuous** function;

U.2 U is is **monotonic**. This means that $q' \succsim q'' \Rightarrow U(q') \geq U(q'')$. To a preferred bundle is associated a higher real number. U is therefore **increasing** along all its margins (positive marginal utilities);

U.3 U is **not unique**. More precisely, it is unique up to any monotonic transformation;

U.4 U is **quasi-concave**. Its level curves – indifference curves – are convex or its upper contour sets are convex.

Remark 1. Strictly speaking, the utility function is non-decreasing (non-negative marginal utilities).

Remark 2. Strictly speaking, the utility function is weakly quasi-concave (non-concave indifference curves).

Remarks 3. We have assumed and we will assume that the utility function is everywhere twice differentiable (we will make the same assumption for all functions we encounter, unless stated otherwise).

2.7 Examples of preferences and utility functions

We have seen that ICs cannot cross each other, are negatively sloped, correspond to higher utility levels the farther away from the origin and are weakly convex. Within this class of preferences, we can highlight a few relevant special cases. We use the two-good case for simplicity.

a. **Perfect substitutes.** Consider a utility function of the form:

$$U = aq_1 + bq_2 \quad \rightarrow -(dq_2/\,dq_1) = a/b \;= MRS_{12} \tag{5}$$

Here a and b are positive scalars referred to as parameters. Note that the MRS, or the slope of a typical IC, is constant, independent of (q_1, q_2). It follows that ICs are straight negatively sloped lines, as shown in the left panel of Figure 2.23. What kind of goods underly these preferences? Suppose for simplicity that $a = b = 1$. The consumer is willing to substitute one good with the other one one-to-one. The two goods are perfectly substitutable. The textbook example are blue and black pens when the agent is only interested in the amount of writing irrespective of the color. But also coffee and tea are perfect substitutes if you are mostly interested in the stimulus they give.

Exercise. What kind of goods are implied by the following utility function: $U = (aq_1 + bq_2)^2$? Are they still perfect substitutes?

b. **Perfect complements.** Consider a utility function of the form:

$$U = \min\{aq_1, bq_2\} \tag{6}$$

Here the agent receives utility from consuming the smallest quantity of one of the two goods combined with any greater quantity of the other good. Put differently, assuming that $a = b = 1$, the agent will not be more satisfied by consuming more than one unit of one good, when consuming one unit of the other one. The typical example are right and left shoes. But also if your tastes are such that you put strictly one spoon of sugar per cup of espresso coffee, or two, or three spoons per single cup but always in fixed proportions. Goods of this sort are said to be perfect complements, as they are always consumed in fixed proportions. ICs of complementary goods have a kink corresponding to the fixed proportions point and are linear along the two (horizontal and vertical) arms. As a consequence, they are not continuously differentiable, so that the MRS is not uniquely defined. ICs of perfect complements are represented on the right panel of Figure 2.23.

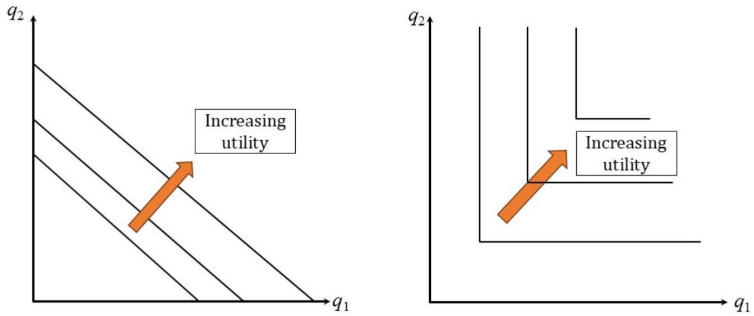

Figure 2.23 *Perfect substitutes and perfect complements.*

c. **Quasi-linear preferences**. Suppose that a consumer has ICs that are vertical translates of one another, implying that all ICs are just vertically shifted upward. These curves represent what are referred to as quasi-linear preferences. They are meant to capture situations where the consumer obtains utility from the quantity of a "minor" good and from the amount of all other goods lumped together. They are not very realistic but have interesting properties that make them popular in welfare analysis. The utility function of quasi-linear preferences has the following form:

$$U = v(q_1) + bq_2 \quad \rightarrow \quad -(dq_2/\,dq_1) = b/\,v'(q_1) \; = MRS_{12}(q_1) \tag{7}$$

Of course, b is positive and so is $v'(q_1)$, being the marginal utilities of the two goods. The ICs of these preferences are protrayed in Figure 2.24.

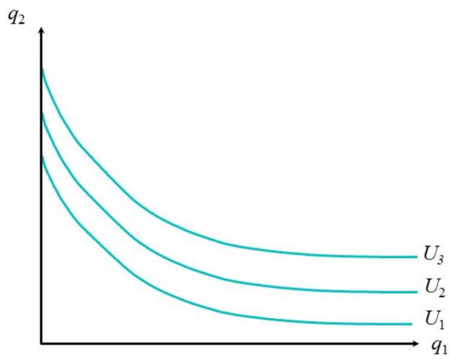

Figure 2.24 *Quasilinear preferences.*

Exercise. Are the utility functions: $U = \sqrt{q_1} + q_2$ and $U = \ln q_1 + q_2$ quasi-linear? Compute the MRS for these two cases.

d. **Cobb-Douglas**. Consider a utility function of the form:

$$U = {q_1}^a {q_2}^b \quad \rightarrow \quad -(dq_2 / \, dq_1) = aq_2 / bq_1 = MRS_{12}\left(q_1, q_2\right) \tag{8}$$

The Cobb-Douglas (CD) functional form was originally proposed and used in production theory, but it proved very convenient both for analytical and didactic purposes also in utility theory. In fact, CD preferences look like the nice convex monotonic ICs that we refer to as "well-behaved" ICs. Depending on the relative value of the two parameters, CD preferences look as in these two forms:

$$U = {q_1}^a {q_2}^b \qquad lnU = alnq_1 + blnq_2 \tag{9}$$

Exercise: Let $U = {q_1}^a {q_2}^b$. Are the transformations $V\left(q_1, q_2\right) = lnU\left(q_1, q_2\right)$ and $V\left(q_1, q_2\right) = U = {q_1}^c {q_2}^{1-c}$ (set $c = \frac{a}{a+b}$) still valid utility functions? Compute the MRS for these cases. What kind of preferences are represented?

2.8 The consumer choice problem

Equipped with objective function and constraint(s), we now proceed to study the agent's optimal choices. The consumer chooses quantities of the various goods so as to maximize her utility subject to the budget constraint (as the goods have to be purchased in the market given prices and her money income). In the two-good case we form the Lagrangian function:

$$\max_{\{q_1, q_{2,\lambda}\}} L = U(q_1, q_2) + \lambda(I - p_1 q_1 - p_2 q_2) \tag{10}$$

where λ is the Lagrange multiplier. We note that, although the quantities of goods can be zero (in which case we should consider Kuhn-Tucker conditions and entertain the possibility of corner solutions), we will proceed by assuming that quantities of goods are strictly positive so as to obtain interior optima.

The optimality conditions entail the calculation of both First order conditions (FOCs) and Second order conditions (SOCs). The FOCs require setting the first derivatives of the Lagrangian with respect to the three choice variables (q_1, q_2, λ) equal to zero. Thus:

$$\begin{cases} L_1 = U_1(q_1^*, q_2^*) - \lambda^* p_1 = 0 \\ L_2 = U_2(q_1^*, q_2^*) - \lambda^* p_2 = 0 \\ L_\lambda = I - p_1 q_1^* - p_2 q_2^* = 0 \end{cases} \tag{11}$$

where asterisks denote the optimal level. To interpret these optimality conditions, we can take the ratio of the first two equations to get:

$$MRS(q_1^*, q_2^*) = \frac{U_1(q_1^*, q_2^*)}{U_2(q_1^*, q_2^*)} = \frac{p_1}{p_2} \tag{12}$$

$$p_1 q_1^* - p_2 q_2^* = I \tag{13}$$

These FOCs tell us two things: (i) the optimal choice must correspond to an expenditure that equals the agent's income, i.e. the optimal bundle must lie on the budget line; (ii) the optimal combination of goods must be such that their marginal rate of substitution is equal to the ratio of their prices. Thus, the agent's subjective trade-off between quantities of goods (dictated by her preferences and the MRS) must equate the market trade-off (dictated by prices). This condition requires the slope of the highest indifference curve to be equal to the slope of the budget line at the optimal point. Graphically:

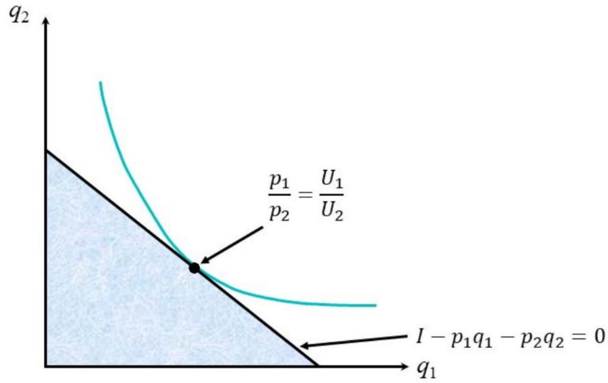

Figure 2.25 *Optimal consumer choice.*

Note that the FOCs are satisfied whether ICs are convex or concave. Consider for a moment concave indifference curves.

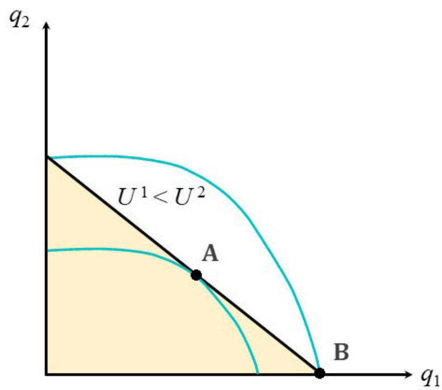

Figure 2.26 *Optimal consumer choice with concave indifference curves.*

It is easy to see that the tangency between indifference curve and budget line leads to a **minimum utility** in A, not a maximum one because it is possible to achieve a higher utility $U^2 > U^1$ without violating the budget constraint. The utility maximizing bundle will thus be a **corner solution** where the agent consumes only one good, as in point B. Although this is a possibility not ruled out by the theory so far, we do not believe it to be representative of the standard choices observed in reality. Our grandmother does not come home from the vegetable market in the street with only celery in her bag, but with quantities of several different vegetables.

To make sure we do not end up with a corner solution or **specialized consumption** we need to make sure that the agent's ICs are convex – to be precise strictly convex – and we have seen that this is guaranteed by the convexity axiom on preferences A.6 above. If we do not want to a priori make this convexity assumption, we need to check the SOCs in order to make sure that we are locating a maximum point, not a minimum one. Note that the strict convexity of the IC ensures uniqueness of the optimal selection. This is ensured analytically by the SOCs to which we now turn.

The SOCs for the problem require differentiating the FOCs system with respect to the same three choice variables and organizing the result in matrix form as follows:

$$H = \begin{bmatrix} U_{11} & U_{12} & -p_1 \\ U_{12} & U_{22} & -p_2 \\ -p_1 & -p_2 & 0 \end{bmatrix} \tag{14}$$

The determinant of this matrix is given by: $|H| = 2U_{12}p_1p_2 - p_2^2U_{11} - p_1^2U_{22}$. For the optimal bundle to be a maximum we must have: $|H| \leq 0$.

Math digression: Checking Second order conditions

Unconstrained optima

In the multivariate case consider the optimization problem:

$$\max_{\{x_1,x_2,\ldots,x_n\}} / \min f\left(x_1, x_2, \ldots, x_n\right)$$

The FOCs or necessary conditions both for a maximum and a minimum are:

$$f_1 = f_2 = \ldots = f_n = 0 \quad \text{where} \quad \partial f / \partial x_i = f_i \quad i = 1, 2, .., n$$

when evaluated at the optimal point $x^* = \left(x_1^*, x_2^*, .., x_n^*\right)$.

The SOCs or sufficient conditions entail the calculation of the Hessian matrix H formed with the second partials of f:

$$H = \begin{bmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{1n} & \cdots & f_{nn} \end{bmatrix} \text{ where } \partial^2 f / \partial x_i \partial x_j = f_{ij} \quad i, j = 1, 2, .., n$$

H has to be negative definite at x^* for a maximum. We have to check the principal minors of H; that is: $H_1 = f_{11} < 0$;

$$H_2 = \begin{vmatrix} f_{11} & f_{12} \\ f_{12} & f_{22} \end{vmatrix} > 0;$$

$$H_3 = \begin{vmatrix} f_{11} & f_{12} & f_{13} \\ f_{12} & f_{22} & f_{23} \\ f_{13} & f_{23} & f_{33} \end{vmatrix} < 0;$$

and so on alternating in sign. Eventually $H_n \gtrless 0$, negative if n is odd and positive if n is even.

Note: we have assumed that cross partials are symmetric. There is no economic reason for that, but this assumption is routinely made in Consumer and Producer theory.

For x^* to be a minimum all principal minors of H have to be positive. Thus:

$$H_1 > 0;$$
$$H_2 > 0;$$
$$H_3 > 0;$$
$$\dots H_n \geq 0.$$

Constrained optima

In the multivariate case consider the optimization problem with one constraint (we will only consider this case as all constrained optimization problems in these notes have a single constraint, though the generalization to many constraints is straightforward):

$$\max_{\{x_1,x_2,\dots,x_n\}} / \min f(x_1, x_2, \dots, x_n) \quad \text{subject to:} \quad g(x_1, x_2, \dots, x_n) = \kappa$$

where κ is a constant. We form the Lagrangian function:

$$\max_{\{x_1,x_2,\dots,x_n,\lambda\}} / \min \mathscr{L}(x_1, x_2, \dots, x_n,) = f(x_1, x_2, \dots, x_n) + \lambda \left[\kappa - g(x_1, x_2, \dots, x_n)\right]$$

where λ is the so-called Lagrange multiplier.

The FOCs or necessary conditions both for a maximum and a minimum require differentiating the Lagrangian with respect to all choice variables and the multiplier and equating them to zero. That is:

$$\frac{\partial \mathscr{L}}{\partial x_i} = f_i + \lambda g_i = 0 \quad \text{for} \quad i = 1, 2, \dots, n$$
$$\frac{\partial \mathscr{L}}{\partial \lambda_i} = \kappa - g(x_1, x_2, \dots, x_n) = 0$$

when evaluated at the optimal point (x^*, λ^*). Note that (i) there are now $(n+1)$ FOCs, (ii) at the optimal point the constraint must always be satisfied with strict equality.

The SOCs or sufficient conditions require first to form the Bordered Hessian matrix B:

$$B = \begin{vmatrix} 0 & g_i \\ g_i & \mathscr{L}_{ij} \end{vmatrix} \quad \text{for} \quad i = 1, 2, \ldots, n$$

This is a $(n+1)$ x $(n+1)$ matrix with the first row and column given by the first partials of the constraint with respect to the choice variable which border the standard Hessian matrix. The principal minors are:

$$B_1 = \begin{vmatrix} 0 & g_1 \\ g_1 & \mathscr{L}_{11} \end{vmatrix} > 0;$$

$$B_2 = \begin{vmatrix} 0 & g_1 & g_2 \\ g_1 & \mathscr{L}_{11} & \mathscr{L}_{12} \\ g_2 & \mathscr{L}_{12} & \mathscr{L}_{22} \end{vmatrix} < 0;$$

and so on.

For a maximum we have to have: $B_1 < 0$; $B_2 > 0$; $B_3 < 0; \ldots; B_n \lesseqgtr 0$ non-positive if n is odd and non-negative if it is even. For a minimum we have to have: $B_1 < 0$; $B_2 < 0$; $B_3 < 0; \ldots; B_n \leq 0$. All minors negative, the last one non-positive.

2.9 Convexity and the second order conditions

The SOCs of our constrained utility maximization problem have an interesting and important implication. The slope of an IC is given by: $-\frac{dq_2}{dq_1} = \frac{U_1(q_1,q_2)}{U_2(q_1,q_2)}$. We want to see how the slope changes as we modify the quantity of one good. That is, we want to study the curvature of the IC. The rate of change of the slope of the IC is $\frac{d\frac{U_1}{U_2}}{dq_1}$. Let us compute the total differential of the numerator:

$$d\frac{U_1}{U_2} = \frac{(U_{11}dq_1 + U_{12}dq_2)U_2 - (U_{21}dq_1 + U_{22}dq_2)U_1}{U_2^2}$$

dividing by dq_1 we get:

$$\frac{d\frac{U_1}{U_2}}{dq_1} = \frac{\left(U_{11} + U_{12}\frac{dq_2}{dq_1}\right)U_2 - \left(U_{21} + U_{22}\frac{dq_2}{dq_1}\right)U_1}{U_2^2}$$

but $-\frac{U_1}{U_2} = \frac{dq_2}{dq_1}$, therefore:

$$\frac{d\frac{U_1}{U_2}}{dq_1} = \frac{U_{11}U_2 - 2U_{12}U_1 + U_{22}\left(\frac{U_1^2}{U_2}\right)}{U_2^2} = (U_2^2U_{11} - 2U_{12}U_1U_2 + U_1^2U_{22})\left(\frac{1}{U_2^3}\right)$$

A convex indifference curve requires $\frac{d\frac{U_1}{U_2}}{dq_1} < 0$. Since $U_2^3 > 0$ and $U_2^2U_{11} - 2U_{12}U_1U_2 + U_1^2U_{22} \leq 0$ by the (weak) quasi-concavity of the utility function, we see that ICs are (weakly) convex. Note that from the FOCs we have $U_1 = \lambda p_1$ and $U_2 = \lambda p_2$. Substituting in the above expression we get:

$$\frac{d\frac{U_1}{U_2}}{dx_1} = \lambda^2(p_2^2U_{11} - 2U_{12}p_1p_2 + p_1^2U_{22})\left(\frac{1}{U_2^3}\right) \tag{15}$$

which must be negative as said. This is the case iff: $(p_2^2U_{11} - 2U_{12}p_1p_2 + p_1^2U_{22}) < 0$ (because λ^2, $U_2^3 > 0$). But this expression is exactly the same as that for the SOCs. We conclude that the convexity of ICs and the SOCs are equivalent for the constrained utility maximization problem. In other words, for the optimal bundle to be a maximum utility bundle we must have convex ICs, which in turn are implied by the axiom of convexity of preferences. This shows the importance of this axiom. Thus:

Axiom Convexity of preferences $\rightarrow$ Quasi-concave utility function $\rightarrow$ Convex indifference curves (convex upper contour set) $\rightarrow$ SOCs satisfied for the constrained utility maximization problem

Remark 1. A diminishing marginal utility (negative second derivative of the utility function) is neither necessary nor sufficient for well-behaved ICs. In Fact $U_{11} < 0$, $U_{22} < 0 \rightarrow p_2^2U_{11} < 0$, $p_1^2U_{22} < 0 \Rightarrow (p_2^2U_{11} - 2U_{12}p_1p_2 + p_1^2U_{22}) < 0$ only if $2U_{12}p_1p_2$ is not positive more than the other terms;

Remark 2. For strictly convex ICs we need to have $|H| = p_2^2U_{11} - 2U_{12}p_1p_2 + p_1^2U_{22} < 0$. If $|H| = 0$ ICs will be linear.

2.10 Marshallian demand functions

Let us go back to the FOCs (11). They form a system of three equations that are implicit functions of the three decisions variables. For many reasons it is convenient to have a set of functional relationships where the choice variables, the optimal quantities of various goods, depend on the problem's exogenous variable, prices and income. This entails solving the system (11) explicitly for (q_1, q_2, λ). To do that we have to invoke the **Implicit function theorem**.

Math digression: Implicit function theorem

Univariate implicit function theorem

Consider an implicit equation $f(x,a) = 0$ and a point (x_0, a_0) solution of the equation so that $f(x_0, a_0) = 0$. Assume:

1. f is continuous and differentiable in a neighborhood of (x_0, a_0);
2. $f_x(x_0, a_0) \neq 0$.

Then: (1) there is one and only one function $x = g(a)$ defined in a neighborhood of a_0 that satisfies $f(g(a),\ a) = 0$ and $x_0 = g(a_0)$; (2) the derivative of $g(a)$ is:

$$g'(a_0) = -\frac{f_a\left[(g(a_0), a_0)\right]}{f_x\left[(g(a_0), a_0)\right]}$$

Multivariate implicit function theorem

Consider a set of equations:

$$\begin{cases} f^1(x_1, .., x_n; a_1, .., a_m) = 0 \\ \qquad \qquad \cdots \\ f^n(x_1, .., x_n; a_1, .., a_m) = 0 \end{cases}$$

and point $(\boldsymbol{x_0}, \boldsymbol{a_0})$ (bold indicates vectors) solution of the equation. Assume: $f^1, \ldots, f^n$ are continuous and differentiable in a neighborhood of $(\boldsymbol{x_0}, \boldsymbol{a_0})$; If the determinant of the following Jacobian matrix is non-zero, $|J| \neq 0$, where:

$$|J| = \begin{vmatrix} \frac{\partial f^1}{\partial x_1} & \cdots & \frac{\partial f^1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f^n}{\partial x_1} & \cdots & \frac{\partial f^n}{\partial x_n} \end{vmatrix}$$

then there exists a system of functions:

$$\begin{cases} x_1 = g^1(a_1, .., a_m) \\ \qquad \qquad \cdots \\ x_n = g^n(a_1, .., a_m) \end{cases}$$

or $\boldsymbol{x} = g(\boldsymbol{a})$ defined in a neighborhood of $\boldsymbol{a_0}$ that satisfies $f(g(\boldsymbol{a}),\ \boldsymbol{a}) = 0$ and $\boldsymbol{x_0} = g(\boldsymbol{a_0})$. Thus: (1) the $(x_1, .., x_n)$ are the dependent variables and $(a_1, .., a_m)$

are the independent variables; (2) the partial derivatives of x_i with respect to a_k at point $(\boldsymbol{x_0}, \boldsymbol{a_0})$ is given by the following:

$$\begin{vmatrix} \frac{\partial x_1}{\partial a_1} & \cdots & \frac{\partial x_1}{\partial a_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial a_1} & \cdots & \frac{\partial x_n}{\partial a_m} \end{vmatrix} = \begin{vmatrix} \frac{\partial f^1}{\partial x_1} & \cdots & \frac{\partial f^1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f^n}{\partial x_1} & \cdots & \frac{\partial f^n}{\partial x_n} \end{vmatrix}^{-1} \begin{vmatrix} \frac{\partial f^1}{\partial a_1} & \cdots & \frac{\partial f^1}{\partial a_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f^n}{\partial a_1} & \cdots & \frac{\partial f^n}{\partial a_m} \end{vmatrix}$$

If the implicit function theorem holds, that is the Jacobian of the FOCs system is non-singular, then we can solve it explicitly for the choice variables as functions of the exogenous variables. Note that the Jacobian of the FOCs (11) is nothing but the SOCs system (14). For a maximum we required $|H| \leq 0$. Now we find out that, to invert the FOCs, it is necessary to strengthen the SOCs to $|H| < 0$. This is equivalent to require the ICs to be strictly convex. From (11) we obtain:

$$\begin{cases} q_1 = \Phi^1(p_1, p_2, I) \\ q_2 = \Phi^2(p_1, p_2, I) \\ \lambda = \Phi^\lambda(p_1, p_2, I) \end{cases} \tag{16}$$

The first two expressions are referred to as **Marshallian or Uncompensated demand functions** for good 1 and good 2. They relate the quantities, which are the endogenous variables of the problem, to the whole set of exogenous variable, prices and income. In general, the quantity demanded for one good will depend on its own and all other prices, besides income.

System (16) defines the **consumer equilibrium**. The structure imposed on the optimization problem, and in particular the curvature of the utility function (strict convexity of ICs), and in turn the convexity axiom on preferences, ensures that such equilibrium exists and is unique.

The last equation of the FOC system refers to the Lagrange multiplier, which is also a function of prices and income.

2.11 Comparative statics

Having characterized the consumer optimal choices and represented them by means of Marshallian demand functions, our next step is to perturb the equilibrium and study the agent's reactions when the economic environment changes. In other words, we wish to induce a change in an exogenous variable, obtain the new equilibrium and compare it with the initial one to see the direction of the change. We will only consider the qualitative (not quantitative, in which case we would need data and econometric methods) change of each one exogeneous variable in turn. This analysis is referred to as **Comparative statics**, as we compare initial and final equilibrium and only see the beginning and the end of the process, not the transition from one equilibrium to the new one (in which case we would have to include time in the picture, making the analysis dynamic).

To perform comparative statics exercises it is convenient to use another mathematical tool which is known as **Cramer's Rule**.

Math digression: Cramer's rule
Suppose we want to determine how the optimal choice functions respond to changes in an exogenous variable or in a parameter a. Start from the FOCs for two goods:

$$\begin{cases} \frac{\partial f[q_1(a),\, q_2(a),\, a]}{\partial q_1} = f_1\left[q_1(a),\, q_2(a),\, a\right] = 0 \\ \frac{\partial f[q_1(a),\, q_2(a),\, a]}{\partial q_2} = f_2\left[q_1(a),\, q_2(a),\, a\right] = 0 \end{cases}$$

Differentiate the two expressions with respect to a:

$$\begin{cases} f_{11}\frac{\partial q_1}{\partial a} + f_{12}\frac{\partial q_2}{\partial a} + f_{13} = 0 \\ f_{12}\frac{\partial q_1}{\partial a} + f_{22}\frac{\partial q_1}{\partial a} + f_{23} = 0 \end{cases}$$

where f_{ij} $(i = q_1, q_2, a)$ are cross partials. Write the above system in compact matrix form:

$$\begin{bmatrix} f_{11} & f_{12} \\ f_{12} & f_{22} \end{bmatrix} \begin{bmatrix} \frac{\partial q_1}{\partial a} \\ \frac{\partial q_2}{\partial a} \end{bmatrix} = \begin{bmatrix} -f_{13} \\ -f_{23} \end{bmatrix}$$

If the matrix on the left-hand side is invertible (its determinant non-singular) we can solve the system to obtain:

$$\begin{bmatrix} \frac{\partial q_1}{\partial a} \\ \frac{\partial q_2}{\partial a} \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} \\ f_{12} & f_{22} \end{bmatrix}^{-1} \begin{bmatrix} -f_{13} \\ -f_{23} \end{bmatrix}$$

This is the response we were looking for. However, there is an often easier method to obtain this result, known as **Cramer's rule**. Given the linear system of equations:

$$\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \quad or \quad Ax = b$$

To find the element x_i of the solution vector of this system of linear equations as a function of a_{ij} and b_i, replace the i-th column of matrix A with the column vector $\boldsymbol{b}$ to form matrix A_i. Then x_i is the determinant of A_i divided by the determinant of A. That is:

$$x_i = \frac{|A_i|}{|A|}$$

We begin by totally differentiating the FOCs:

$$\begin{cases} U_{11}dq_1 + U_{12}dq_2 - p_1 d\lambda - \lambda dp_1 = 0 \\ U_{21}dq_1 + U_{22}dq_2 - p_2 d\lambda - \lambda dp_2 = 0 \\ -p_1 dq_1 - p_2 dq_2 + dI - q_1 dp_1 - q_2 dp_2 = 0 \end{cases} \tag{17}$$

or in matrix form:

$$\begin{bmatrix} U_{11} & U_{12} & -p_1 \\ U_{12} & U_{22} & -p_2 \\ -p_1 & -p_2 & 0 \end{bmatrix} \begin{bmatrix} dq_1 \\ dq_2 \\ d\lambda \end{bmatrix} = \begin{bmatrix} \lambda dp_1 \\ \lambda dp_2 \\ -dI + q_1 dp_1 + q_2 dp_2 \end{bmatrix} \tag{18}$$

Income effect. We now consider the change in the demand for good 1, q_1, brought about by a change in money income I. Note that the matrix on the left-hand side of (18) is nothing but the SOCs of our problem. To compute $\frac{dq_1}{dI}$ we let $dI \neq 0$ and $dp_1 = dp_2 = 0$. Thus:

$$\frac{dq_1}{dI} = \frac{\begin{vmatrix} 0 & U_{12} & -p_1 \\ 0 & U_{22} & -p_2 \\ -1 & -p_2 & 0 \end{vmatrix}}{|H|} \tag{19}$$

which yields:

$$\frac{dq_1}{dI} = \frac{-(p_1 U_{22} - p_2 U_{12})}{|H|} \lesseqgtr 0 \tag{20}$$

A first result we obtain is that an increase (resp. decrease) in money income does not necessarily lead the consumer to increase (resp. decrease) the demand for good q_1, since the sign of (20) depends on the sign of the numerator $(p_1 U_{22} - p_2 U_{12}) \gtrless 0$. Indeed, we define:

Normal good. Good 1 is said to be normal if $\left.\frac{dq_1}{dI}\right|_{dp_i=0} > 0$;

Inferior good. Good 1 is said to be inferior if $\left.\frac{dq_1}{dI}\right|_{dp_i=0} < 0$.

Most goods are normal, but some can be inferior, e.g. margarine (relative to butter).

Graphically, an increase in income shifts the budget line outward as a parallel new line (prices have not changed). The new optimal bundle again satisfies the tangency condition for optimality and contains different quantities of the two goods. The quantity of good 1 increases as it is a normal good. However, the quantity of good 2 decreases, meaning that good 2 is inferior. Note that the agent's utility goes up. See Figure 2.27.

Own price effect. We now consider the change in the demand for good 1, q_1, brought about by a change in its own price p_1. To compute $\frac{dq_1}{dp_1}$ we let $dp_1 \neq 0$ and $dI = dp_2 = 0$. Thus:

$$\frac{dq_1}{dp_1} = \frac{\begin{vmatrix} \lambda & U_{12} & -p_1 \\ 0 & U_{22} & -p_2 \\ q_1 & -p_2 & 0 \end{vmatrix}}{|H|} \tag{21}$$

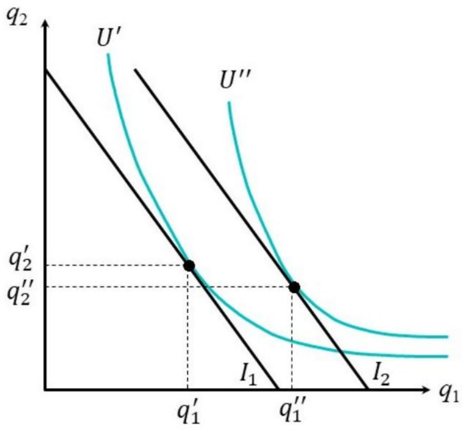

Figure 2.27 *Increase in income and normal and inferior goods.*

which yields:

$$\frac{dq_1}{dp_1} = \frac{-q_1 p_2 U_{12} + q_1 p_1 U_{22} - p_2^2 \lambda}{|H|} = -\frac{p_2^2 \lambda}{|H|} - q_1 \frac{p_2 U_{12} - p_1 U_{22}}{|H|} \lesseqgtr 0 \tag{22}$$

Also in this case the effect is ambiguous. Contrary to usual expectations, an increase in its own price may not induce a decrease in the demand for a good. Let's see more precisely the origin of this ambiguity.

Note from (22) that the first term is negative as the components of the ratio are all positive. The ambiguity of the sign in entirely due to the numerator of the second term. But from (20) that term is easily recognizable as the income effect. We conclude that:

a. the own price effect is negative $\frac{dq_1}{dp_1} < 0$ if the good is normal;

b. the own price effect is ambiguous $\frac{dq_1}{dp_1} \gtrless 0$ if the good is inferior, as we do not know which one of the two terms of (22) prevails.

We see that, as most goods are normal, their demand increases (resp. decreases) if the own price goes down (resp. up). But the theory does not deliver automatically such result.

Exercise: Compute the own price effect $\frac{dq_2}{dp_2}$ and study its sign.

Cross price effect. We now consider the change in the demand for good 1, q_1, brought about by a change in the other price p_2. To compute $\frac{dq_1}{dp_2}$ we let $dp_2 \neq 0$ and $dI = dp_1 = 0$. Thus:

$$\frac{dq_1}{dp_2} = \frac{\begin{vmatrix} \lambda & U_{12} & -p_1 \\ 0 & U_{22} & -p_2 \\ q_1 & -p_2 & 0 \end{vmatrix}}{|H|} \tag{23}$$

which yields:

$$\frac{dq_1}{dp_2} = \frac{p_1 p_2 \lambda}{|H|} - q_2 \frac{p_2 U_{12} - p_1 U_{22}}{|H|} \lesseqgtr 0 \tag{24}$$

Again the sign is ambiguous.

Exercise: Compute the cross price effect $\frac{dq_2}{dp_1}$ and study its sign.

2.12 Engel and demand curves

We now investigate graphically the impact of changes in the exogenous variables just analyzed. Increases in income shift the budget line outward in a parallel way. This leads to new optimal bundles which include larger quantities of the two goods, assumed to be normal, and a higher maximum utility. This is shown in Figure 2.28. If we connect the optimal points, we trace a curve that is referred to as **income expansion path** or **income-consumption curve**.

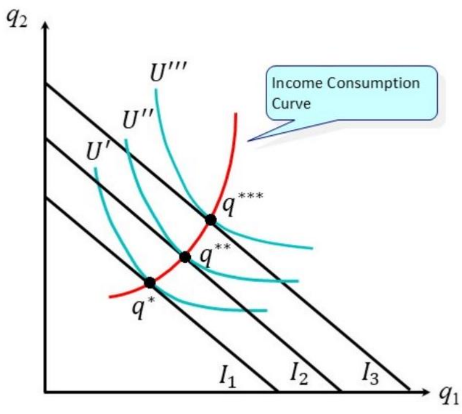

Figure 2.28 *Increase in income and income-consumption curve.*

We see that the curve is positively sloped because the two goods are normal goods. With one inferior good the curve would slope downward.

Let us now focus on the demand for good 1 and project the optimal quantities of q_1 and the corresponding levels of income I onto the new space (q_1, I), so as to obtain the following Figure 2.29.

The curve that is obtained is referred to as **Engel curve**. It is easily seen that the Engel curve slopes upward (resp. downward) for normal (resp. inferior) goods. The expression of the Engel curve is: $q_1 = \varphi(I, p_1^0, p_2^0)$.

Turning to the (own) price effect, consider several decreases of the price of good 1. We know that the budget line rotates around the intercept with the q_2 axis (whose price has not changed) becoming flatter and flatter. We see that as the price p_1 goes down (resp. up), the quantity q_1 increases (resp. decreases). If we connect

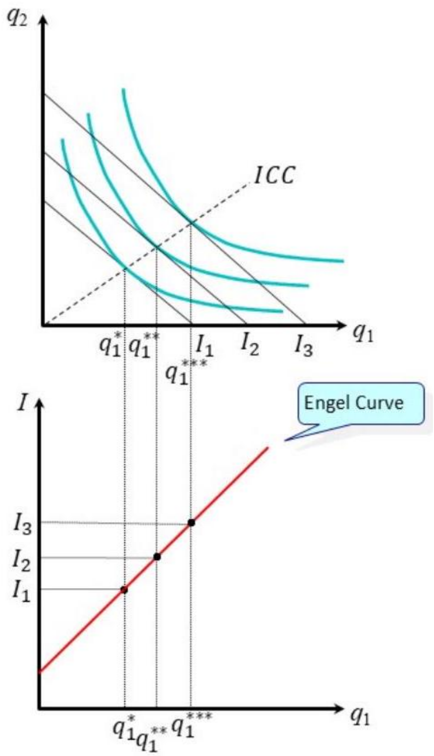

Figure 2.29 *Increase in income and the Engel curve.*

the optimal points so obtained, we trace the so-called **price-consumption curve**, portrayed in Figure 2.30.

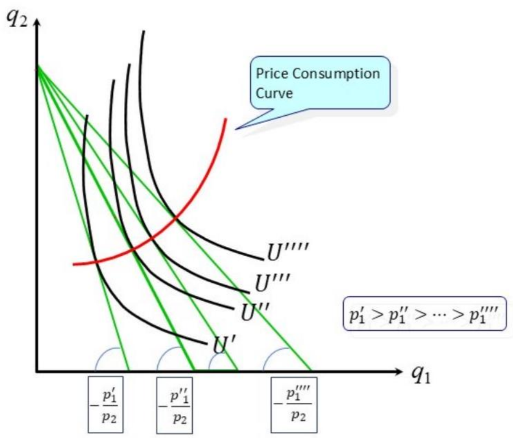

Figure 2.30 *Decrease in price and price-consumption curve.*

If, as before, we project the optimal quantities of q_1 and the corresponding levels of price p_1 onto the new space (q_1, p_1), we get the following situation (Figure 2.31).

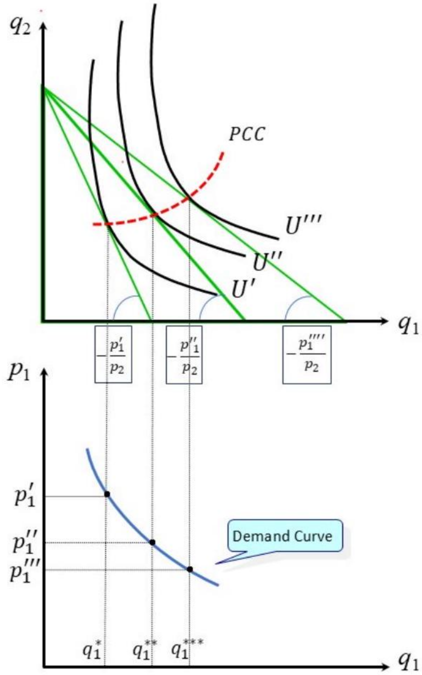

Figure 2.31 *Increase in price and the demand curve.*

We see that what we obtain is nothing but the (Marshallian) demand curve for good 1, represented for given price of good 2 and income levels.

An interesting situation arises when preferences produce linear income expansion paths. This implies that the slope of indifference curves is constant along rays stemming from the origin, so that the Engel curve for each good is linear. Situations such as these are associated to **Homothetic utility functions**. In consumer theory a homothetic utility function is a monotonic transformation of a function which is homogeneous of degree 1, That is:

$$V(q_1, q_2) = H[U(q_1, q_2)]$$

where $H' > 0$ and $tU(q_1, q_2) = U(tq_1, tq_2)$. Alternatively, if $U(tq_1, tq_2) = g(t)U(q_1, q_2)$ for $t > 0$, then the function U is homothetic.

A relevant property of homothetic functions is that the ratio between first partials (marginal utilities) is independent of scale. That is: $MRS_{12}(q_1, q_2) = MRS_{12}(tq_1, tq_2)$. In fact:

$$\frac{U_1(tq_1, tq_2)}{U_2(tq_1, tq_2)} = \frac{g(t)U_1(q_1, q_2)}{g(t)U_2(q_1, q_2)} = \frac{U_1(q_1, q_2)}{U_2(q_1, q_2)}$$

While homothetic functions are relevant in consumer theory, they are much less in producer theory, where homogenous functions play a much more important role.

Math digression: Homogeneous and Homothetic functions
A function $f(\underline{x})$ is homothetic if:

$$f(\underline{x}) = F(\varphi(\underline{x}))$$

where $F' > 0$ and $\varphi(\underline{x})$ is homogeneous of degree one (i.e. linearly homogeneous). Note that the homothetic function is a monotonic transformation of a homogeneous one, but **not all** homothetic functions are also homogeneous. **All** homogeneous functions are instead also homothetic (putting $F' = 1$).

Lemma:
If $f(\underline{x})$ is homothetic then:

$$\frac{\frac{\partial f(\beta \underline{x})}{\partial x_i}}{\frac{\partial f(\beta \underline{x})}{\partial x_j}} = \frac{\frac{\partial \varphi(\underline{x})}{\partial x_i}}{\frac{\partial \varphi(\underline{x})}{\partial x_j}}$$

Proof. If $f(\underline{x})$ is homothetic then:

$$\frac{\partial f(\beta \underline{x})}{\partial x_i} = \frac{F' \partial \varphi(\beta \underline{x})}{\partial x_i} = F' \frac{\partial \varphi(\underline{x})}{\partial x_i}$$

as $\varphi(\underline{x})$ is linear homogenous and its first partial if homogenous of degree zero. it follows that:

$$\frac{\frac{\partial f(\beta \underline{x})}{\partial x_i}}{\frac{\partial f(\beta \underline{x})}{\partial x_j}} = \frac{\frac{\partial \varphi(\underline{x})}{\partial x_i}}{\frac{\partial \varphi(\underline{x})}{\partial x_j}}$$

What is the economic meaning? Consider the income expansion path in Figure 2.32.

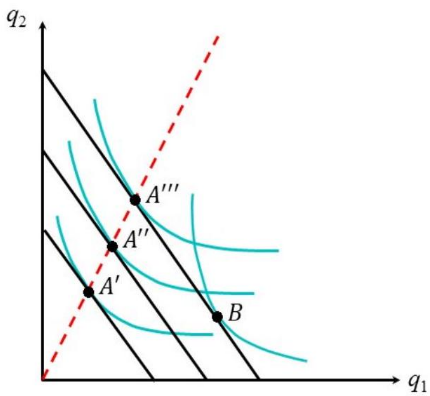

Figure 2.32 *Income expansion path of homothetic utility functions.*

It is the locus of equilibrium choices where the marginal rate of substitution between any two goods is equal to the corresponding relative prices: $\frac{U_1}{U_2} = \frac{p_1}{p_2}$. If the utility function is homothetic then:

$$\frac{U_1}{U_2} = \frac{U_1(\beta x_1)}{U_2(\beta x_2)} = \frac{p_1}{p_2}$$

which means that the new equilibrium lies on a straight line stemming from the origin. In the new equilibrium one has the same proportion of goods $\frac{q_1^*}{q_2^*}$, i.e. the same consumption ratio. We conclude that **homothetic utility functions generate linear income expansion paths**.

For future references we note that the indirect utility function – see below – can be written as a linear function of income $V(p_1, p_2, I) = v(p_1, p_2) I$.

2.13 Optimal choice with perfect substitutes and perfect complements

In the case of perfect substitutes, the optimal choice is not characterized by a unique interior optimum. Figure 2.33 shows that we have three possible cases (in the figure the MRS is equal to – 1 for simplicity).

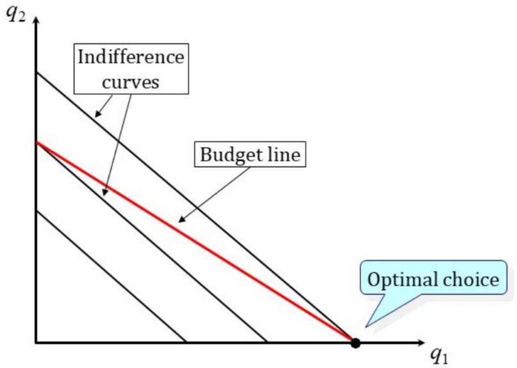

Figure 2.33 *Optimal choice with perfect substitutable goods.*

If $p_2 > p_1$ the slope of the budget line is flatter than the slope of the ICs. In this case, the optimal bundle is where the consumer spends all her money income on good 1. It is a **corner solution**. If, on the contrary, $p_1 > p_2$, then the consumer purchases only good 2, another corner solution. Finally, if $p_1 = p_2$ there is a whole range of optimal choices — any amount of goods 1 and 2 that satisfies the budget constraint is optimal in this case. There are **multiple equilibria**. The demand function for the perfect substitutable good 1 will be:

$$q_1 = \begin{cases} I/p_1 \;\; for \; p_1 < p_2 \\ \text{any value between 0 and } I/p_1 \text{ for } p_1 = p_2 \\ 0 \;\; for \; p_1 > p_2 \end{cases}$$

The textbook example of these goods is black and blue pen when utility is derived by the amount of writing regardless of color. I will then buy only blue or black pens depending on which is cheaper. If two goods are perfect substitutes, then a consumer will purchase the cheaper one. If both goods have the same price, then the consumer doesn't care which one she purchases.

When there is a price change the consumer will jump from one corner solution to the opposite one. The price change is entirely due to a substitution effect. This is shown in Figure 2.34.

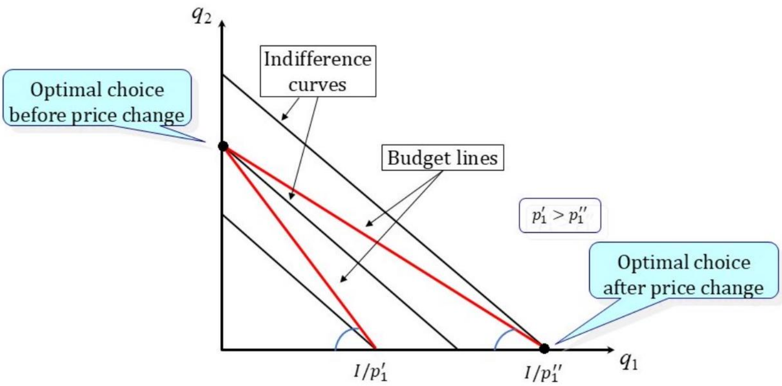

Figure 2.34 *Price change with perfect substitutable goods.*

The polar case of perfect complementary goods is illustrated in Figure 2.35.

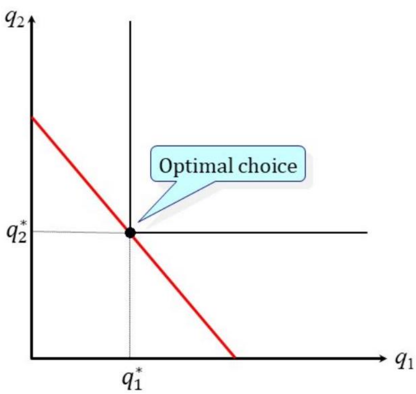

Figure 2.35 *Optimal choice with perfect complementary goods.*

Note that the optimal choice must always lie where the budget line touches the highest indifference curve at the kink, where the consumer is purchasing equal amounts (or amounts in fixed proportions) of both goods, no matter what the prices are. The stupidest example notes that people with two feet buy shoes in pairs. Suppose the consumer purchases the same amount q of good 1 and of good 2, no matter what the prices are. This has to satisfy the budget constraint $p_1 q_1 + p_2 q_2 = I$. Solving for q yields the optimal choices of goods 1 and 2: $q_1 = q_2 = \frac{I}{p_1+p_2}$. This is the demand function for the composite good: since the two goods are always consumed together, it is just as if the consumer were spending all of her money on a single good that had a price of $p_1 + p_2$. It is important to note that relative prices play no role with perfect complements, as optimal decisions are not affected by changes in relative prices. In Figure 2.36 a change in relative prices induces no substitution between the two goods as shown by the budget line that is still tangent to the initial equilibrium point but reflects final prices. The new equilibrium reflects only the income effect.

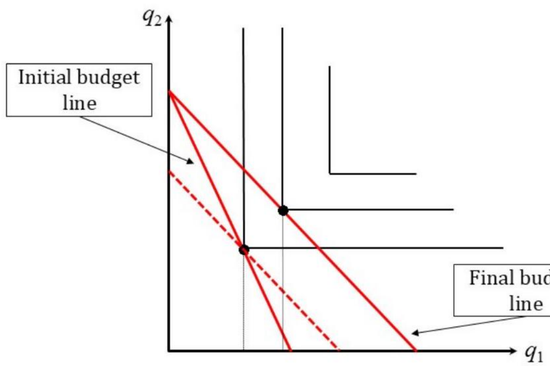

Figure 2.36 *Price change with perfect complementary goods.*

2.14 The indirect utility function and the economic interpretation of the Lagrange multiplier

If we substitute the system of Marshallian demand functions (16) into the objective function (10) we obtain the optimal value function which is referred to as **Indirect Utility Function** (IUF). This function defines the maximum utility attainable by the consumer for given prices and income. The IUF therefore depends on the problem's exogenous variables. Thus:

$$V(p_1, p_2, I) = U[q_1(p_1, p_2, I), q_2(p_1, p_2, I)] \tag{25}$$

The Lagrange multiplier of the constrained utility maximization problem has an interesting economic interpretation. Generally speaking, the multipliers of all the constrained optimization problems of Microeconomics represent so-called shadow prices, i.e. non-market values of prices associated with the constraint they refer to. In the present case, to show the economic interpretation of this Lagrange multiplier we need to be familiar with a mathematical result known as the **Envelope theorem**.

Math digression: Envelope theorem

1. Unconstrained optimization

Let $\boldsymbol{x}(a) = \underset{\boldsymbol{x}}{\operatorname{argmax}} f(\boldsymbol{x}, a)$ be the optimal solution of an unconstrained maximization problem, where a is a parameter. Similar considerations hold for a minimum. We can define the (optimal) **value function** $V(a) = f[\boldsymbol{x}(a), a]$. This tells us the optimized value of the objective function for alternative values of a.

We are often interested in how the optimized value changes as parameter a changes. Clearly a change in a affects V both directly and through its effect on $f(a)$. That is:

$$\frac{dV(a)}{da} = \frac{\partial V(a)}{\partial \boldsymbol{x}}\frac{\partial \boldsymbol{x}(a)}{\partial a} + \frac{\partial V(a)}{\partial a} = \frac{\partial f[\boldsymbol{x}(a), a]}{\partial \boldsymbol{x}}\frac{\partial \boldsymbol{x}(a)}{\partial a} + \frac{\partial f[\boldsymbol{x}(a), (a)]}{\partial a}$$

But because the FOCs for a maximum imply that $\frac{\partial f[x(a), a]}{\partial \boldsymbol{x}} = 0$ then we have that the above change reduces to:

$$\frac{dV(a)}{da} = \frac{\partial V(a)}{\partial a} = \frac{\partial f[\boldsymbol{x}(a), (a)]}{\partial a}$$

Simply put, the Envelope theorem states that the indirect impact of a on V through f is negligible, or zero. Only the direct effect matters.

2. Constrained optimization

Consider the constrained maximization problem: $\boldsymbol{x}(a) = \underset{\boldsymbol{x}}{\operatorname{argmax}} \left[f(\boldsymbol{x}, a) \,|\, h(\boldsymbol{x}, a)\right]$.

The Lagrangian is:

$$\max_{\{\boldsymbol{x}, \lambda\}} \mathscr{L} = f(\boldsymbol{x}, a) - \lambda g(\boldsymbol{x}, a)$$

And the FOCs are:

$$\begin{cases} \frac{\partial f(\boldsymbol{x}, a)}{\partial \boldsymbol{x}} - \lambda \frac{\partial g(\boldsymbol{x}, a)}{\partial \boldsymbol{x}} = 0 \\ g(\boldsymbol{x}, a) = 0 \end{cases}$$

They yield the optimal rules $\boldsymbol{x} = \boldsymbol{x}(a)$ and $\lambda = \lambda(a)$ and the value function $V(a) = f[x(a), (a)]$. We note that: $V(a) = \mathscr{L}(a) = f[\boldsymbol{x}(a), a] - \lambda(a) g[\boldsymbol{x}(a), a)] = f[\boldsymbol{x}(a), a]$ because from the FOCs above we have that $g(\boldsymbol{x}, a) = 0$. Thus, the optimized Lagrangian is the same as the optimal value function.

Now consider a change in parameter a. We have that:

$$\frac{dV(a)}{da} = \frac{d\mathscr{L}(a)}{da} =$$
$$= \left[\frac{\partial f\left[\boldsymbol{x}(a)\right]}{\partial \boldsymbol{x}} - \lambda \frac{\partial g\left(\boldsymbol{x}, a\right)}{\partial \boldsymbol{x}}\right] \frac{\partial \boldsymbol{x}(a)}{\partial a} + \frac{\partial f\left[\boldsymbol{x}(a)\right]}{\partial a} - \frac{\partial \lambda(a)}{\partial a} g\left[\boldsymbol{x}(a), a\right)] - \lambda \frac{\partial g(a)}{\partial a}$$

As before, according to the Envelope theorem only the direct effects matter. In the above expression the indirect effects are zero because of the FOCs, so that $\frac{\partial f[\boldsymbol{x}(a)]}{\partial \boldsymbol{x}} - \lambda \frac{\partial g(\boldsymbol{x},a)}{\partial \boldsymbol{x}} = 0$ and $g\left[\boldsymbol{x}(a), a\right)] = 0$. In the end we are only left with the direct effect:

$$\frac{dV(a)}{da} = \frac{\partial f\left[\boldsymbol{x}(a)\right]}{\partial a} - \lambda \frac{\partial g(a)}{\partial a}$$

Let us now apply the Envelope theorem to our problem. We ask what is the impact of a change in income on the IUF. To that end we apply the Envelope theorem. Using the optimized Lagrangian (10) we have:

$$\frac{dV\left(p_1, p_2, I\right)}{dI} = \frac{d\mathscr{L}\left(p_1, p_2, I\right)}{dI} =$$
$$= \left[\frac{\partial U\left(q_1, q_2\right)}{\partial q_1} - \lambda p_1\right] \frac{\partial q_1}{\partial I} + \left[\frac{\partial U\left(q_1, q_2\right)}{\partial q_2} - \lambda p_2\right] \frac{\partial q_2}{dI} + \frac{\partial \lambda}{\partial I}\left(I - p_1 q_1 - p_2 q_2\right) + \lambda$$

The FOCs imply that the first three terms of the above expression are zero, so that the final result is:

$$\frac{dV\left(p_1, p_2, I\right)}{dI} = \lambda \tag{26}$$

The Lagrange multiplier of the consumer constrained utilization problem is the **marginal utility of income**: it represents how the maximum utility changes if there is a change in money income.

2.15 Expenditure minimization and Hicksian demand functions

Consumer choices can also be analyzed from a different, related perspective. Rather than selecting the bundle that maximizes her utility conditional of income, we could ask what is the bundle that minimizes the total expenditure that enables her to attain a given utility level. Graphically, rather than achieving the highest indifference curve compatible with the given budget line, the problem becomes that of selecting the lowest budget/expenditure line compatible with the given indifference curve. This is shown in Figure 2.37.

The problem is to $\min\limits_{\{q_1, q_2\}} p_1 q_1 + p_2 q_2$ subject to $U(q_1, q_2) = U^0$. We form the Lagrange function:

$$\min_{\{q_1,\ q_2,\ \mu\}} \mathscr{L} = p_1 q_1 + p_2 q_2 + \mu\left[U^0 - U(q_1, q_2)\right] \tag{27}$$

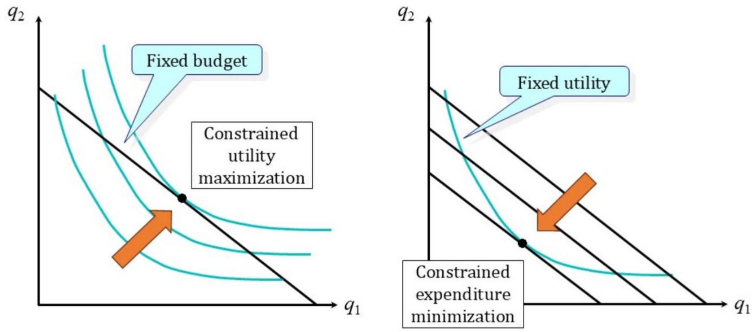

Figure 2.37 *Utility maximization and expenditure minimization.*

The FOCs are:

$$\begin{cases} \mathscr{L}_1 = p_1 - \mu U_1(q_1^*, q_2^*) = 0 \\ \mathscr{L}_2 = p_2 - \mu U_2(q_1^*, q_2^*) = 0 \\ \mathscr{L}_\mu = U^0 - U(q_1^*, q_2^*) = 0 \end{cases} \tag{28}$$

where asterisks denote the optimal quantities. Take the ratio of the first two equations to get:

$$MRS\left(q_1^*, q_2^*\right) = \frac{U_1(q_1^*, q_2^*)}{U_2(q_1^*, q_2^*)} = \frac{p_1}{p_2} \tag{29}$$

$$U^0 = U(q_1^*, q_2^*) \tag{30}$$

According to the FOC (30) the optimal bundle must lie on the given IC. Moreover, interestingly we see that the optimal combination of goods must be such that their marginal rate of substitution is equal to the ratio of their prices: see (29). This is the same condition we obtained in the constrained utility maximization problem – see equation (10) – so that Figure 2.25 applies here as well. The SOCs of the problem require the matrix:

$$A = \begin{bmatrix} -\mu U_{11} & -\mu U_{12} & -U_1 \\ -\mu U_{12} & -\mu U_{22} & -U_2 \\ -U_1 & -U_2 & 0 \end{bmatrix}$$

to be negative semi-definite. It is worth computing the determinant of A to get:

$$|A| = \mu(U_1^2 U_{22} - 2U_1 U_2 U_{22} + U_2^2 U_{11}) \leq 0 \tag{31}$$

But this expression is nothing but the Bordered Hessian of the utility function and is necessarily negative (non-negative) by its assumed (weak) quasi-concavity (and in turn by the convexity axiom on preferences).

We now restrict (31) to be strictly negative – or the utility function to be strictly q-concave – so that we are guaranteed that the IFT holds and the FOCs system (28) can be solved for the optimal quantities:

$$\begin{cases} q_1 = \phi^1(p_1, p_2, U^0) \\ q_2 = \phi^2(p_1, p_2, U^0) \\ \mu = \phi^\mu(p_1, p_2, U^0) \end{cases} \tag{32}$$

These functions are referred to as **Hicksian or Compensated demand functions**. They relate the optimal quantities of the two goods to prices and utility levels.

We also here conduct a few comparative statics exercises. Specifically, we are interested in price effects. Totally differentiating the FOCs and rearranging:

$$[A]\begin{bmatrix} dq_1 \\ dq_2 \\ d\mu \end{bmatrix} = \begin{bmatrix} -dp_1 \\ -dp_2 \\ -dU^0 \end{bmatrix} \tag{33}$$

Own price effect. Let $dp_1 \neq 0$ and $dp_2 = dU^0 = 0$ and solve for $\frac{dq_1}{dp_1}$ to get:

$$\frac{dq_1}{dp_1} = \frac{\begin{vmatrix} -1 & -\mu U_{11} & -U_1 \\ 0 & -\mu U_{22} & -U_2 \\ 0 & -U_2 & 0 \end{vmatrix}}{|A|} = \frac{U_2^2}{|A|} < 0 \tag{34}$$

We see that unambiguously Hicksian demand functions are negatively related to the own price. It can be shown that also $\frac{dq_2}{dp_2} < 0$, so that the own price effect is always negative. Hence, **Hicksian demand curves always slope downward** (unlike Marshallian demand curves). Graphically, the situation is the following:

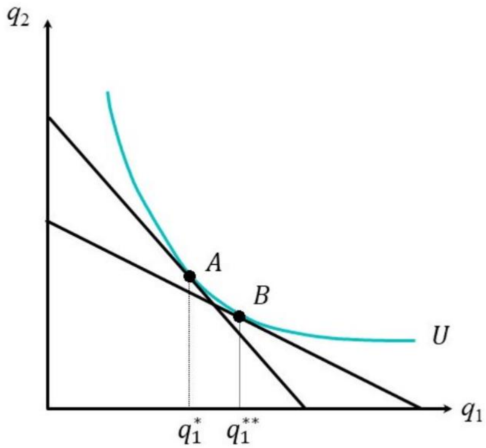

Figure 2.38 *Hicksian price effect.*

We see that the price change leads the isocost line to change slope and the equilibrium point to move along the given indifference curve. The isocost line becomes less steep if the price of good 1 decreases thereby inducing an increase in the quantity demanded. It is important to note that the negative slope of ICs is responsible for the negative price effect. As the agent increases the quantity of good 1, now cheaper, at the same time reducing the quantity of good 2 now more expensive, she effectively substitutes the consumption of the two goods. Hence the price effect is referred to as the pure **substitution effect**, which is unambiguously negative.

Cross price effect. Let $dp_2 \neq 0$ and $dp_1 = dU^0 = 0$ and solve for $\frac{dq_1}{dp_2}$ to get:

$$\frac{dq_1}{dp_2} = \frac{\begin{vmatrix} 0 & -\mu U_{11} & -U_1 \\ -1 & -\mu U_{22} & -U_2 \\ 0 & -U_2 & 0 \end{vmatrix}}{|A|} = -\frac{U_1 U_2}{|A|} > 0 \tag{35}$$

This effect is unambiguously positive, meaning that when the price of good 2 goes up, and its demand goes down, the demand for good 1 goes up.

An interesting additional fact is that $\frac{dq_2}{dp_1} = -\frac{U_1 U_2}{|A|} > 0$, so that the cross price effects of Hicksian demand functions are symmetric: $\frac{dq_1}{dp_2} = \frac{dq_2}{dp_1}$.

More generally, for every pair of n goods, the cross price effect is such that:

1. If the cross price effect is positive $\frac{dq_i}{dp_j} > 0$ the two goods are **substitutes**;
2. If the cross price effect is negative $\frac{dq_i}{dp_j} < 0$ the two goods are **complements**;
3. If the cross price effect is zero $\frac{dq_i}{dp_j} = 0$ the two goods are **independent**.

We note that in a two-good world the two goods can only be substitutes. With more than two goods, holding the quantity of all other goods constant, we may have cross price effects for any pair of goods to be either positive or negative (or equal to zero).

2.16 Slutsky equation

Recall from (34) that $\left.\frac{dq_1}{dp_1}\right|_{dU=0} = \frac{U_2^2}{|A|} < 0$ is the **negative substitution effect**. Let us now consider the relationship between the expenditure minimization problem and the utility maximization problem, or the relationship between Marshallian and Hicksian demand functions. We are interested in the (own) price effects.

Recall the FOCs (11):

$$U_i(.) = \lambda p_i \quad \Longrightarrow \quad U_i / p_i = \lambda \qquad i = 1,2$$

Recall the FOCs (28):

$$\mu U_i(.) = p_i \quad \Longrightarrow \quad p_i / U_i = \mu \qquad i = 1,2$$

It follows that:

$$\mu = \frac{1}{\lambda} \tag{36}$$

Thus, the Lagrange multipliers are one the reciprocal of the other. Consider now the SOCs of the two problems and note that $p_i = U_i/\lambda$:

$$|H| = p_2^2 U_{11} - 2U_{12} p_1 p_2 + p_1^2 U_{22} = \frac{1}{\lambda^2}\left(U_2^2 U_{11} - 2U_{12} U_1 U_2 + U_1^2 U_{22}\right)$$

and:

$$|A| = \mu(U_1^2 U_{22} - 2U_1 U_2 U_{22} + U_2^2 U_{11})$$

It follows that:

$$|A| = \lambda |H| \tag{37}$$

which shows how also the SOCs of the two problems are related.

Note that we can use the FOCs $U_i(.) = \lambda p_i$ and (35) to express the Hicksian own price effect as follows:

$$\left.\frac{dq_1}{dp_1}\right|_{dU=0} = \frac{U_2^2}{|A|} = -\frac{\lambda^2 p_2^2}{\lambda |H|} = -\frac{\lambda p_2^2}{|H|} < 0 \tag{38}$$

Let us go back to the Marshallian own price effect (22):

$$\frac{dq_1}{dp_1} = \frac{-q_1 p_2 U_{12} + q_1 p_1 U_{22} - p_2^2 \lambda}{|H|} = -\frac{p_2^2 \lambda}{|H|} - q_1 \frac{p_2 U_{12} - p_1 U_{22}}{|H|} \lesseqgtr 0 \tag{39}$$

We see that the first term of the above expression is nothing but the (Hicksian) substitution effect (38) and the second term is the income effect (20) so that we may rewrite it as follows:

$$\frac{dq_1}{dp_1} = \left(\frac{dq_1}{dp_1}\right)_{dU=0} - q_1 \left(\frac{dq_1}{dI}\right)_{dp=0} \tag{40}$$

This is known as the **Slutsky equation** and decomposes the total (Marshallian) price effect into the algebraic sum of the Hicksian price or substitution effect and of the income effect. As a further elaboration, since from the budget constraint $dI = q_i dp_i$, (38) can also be represented as follows:

$$\frac{dq_1}{dp_1} = \left(\frac{\partial q_1}{\partial p_1}\right)_{dU=0} - \left(\frac{\partial I}{\partial p_1}\right)_{dp_2=0} \left(\frac{\partial q_1}{\partial I}\right)_{dp=0} \tag{41}$$

Let us see this decomposition graphically. Suppose that $dp_1 < 0$ and $dp_2 = dI = 0$. The budget constraint rotates and becomes flatter and $q_1^C - q_1^A$ is the total effect of $\frac{dq_1}{dp_1}$. That is:

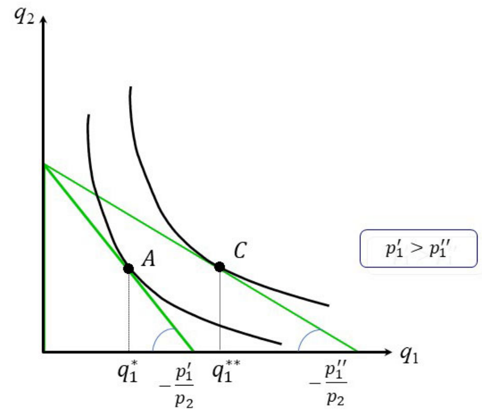

Figure 2.39 *Total price effect.*

This total effect can be broken in two pieces:

1. $q_1^B - q_1^A$ is the substitution (compensated) effect on dq_1 in the sense that $dU = 0$. When the price changes we suppose that the consumer first reacts by substituting the good which has become more expensive with the other one, which is now cheaper, maintaining the same level of utility initially enjoyed. Thus, as a first step there is a substitution between the two goods represented by a movement along the initial indifference curve. This is what happens for a price change within the Hicksian demand framework. Graphically:

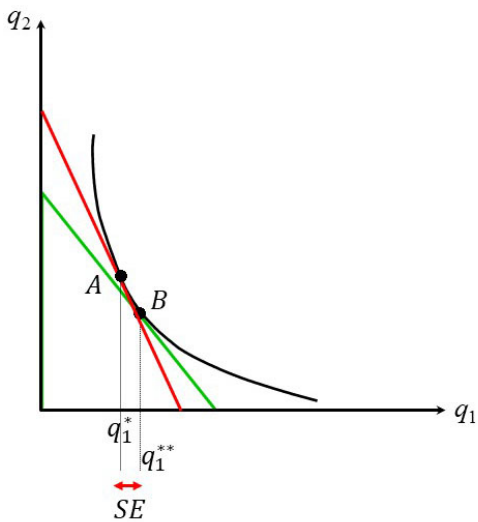

Figure 2.40 *Substitution effect.*

2. $q_1^C - q_1^B$ is the income effect. Right after this change the agent realizes that the price change has altered the purchasing power of her money income, which then induces a further change in the quantity of good 1 – actually, of both goods – due to an income effect. The change in purchasing power of income is equivalent to a parallel shift of the budget line: it is an income effect. Graphically:

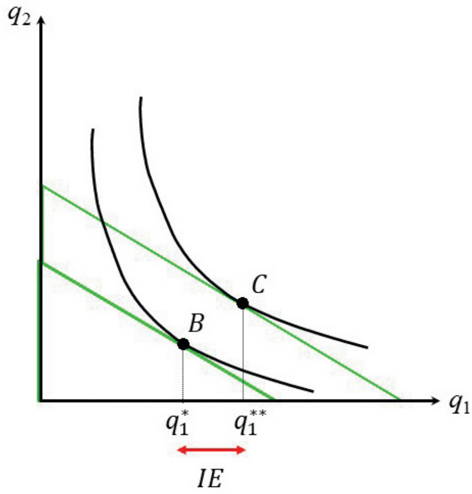

Figure 2.41 *Income effect.*

Thus, the total price effect is the algebraic sum of a substitution effect and an income effect. Putting the two effects together we have:

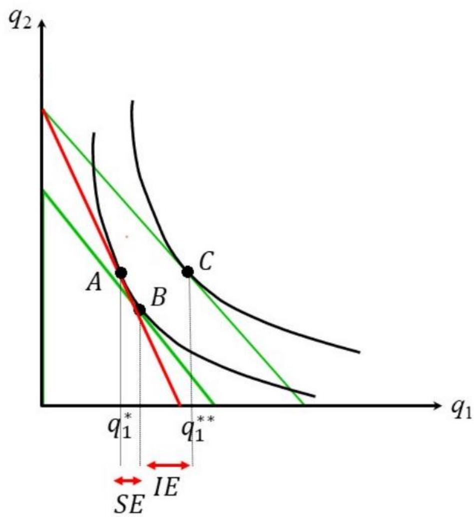

Figure 2.42 *Slutsky decomposition.*

Note from (38) that:

1. If the good is normal the substitution effect is negative and the income effect is positive: there is no conflict between the two terms of (38) and the total price effect is negative. The **Marshallian demand curve is negatively sloped**.

2. If the good is inferior the substitution effect is negative and so is the income effect: this makes the second term of (38) positive and there is conflict between the two terms of the equation. If the second term does not prevail the **Marshallian demand curve remains negatively sloped**.

3. If the good is inferior and the negative income effect is so strong as to offset the substitution effect equation (38) becomes positive, so that the **Marshallian demand curve slopes upward**. If this is the case the inferior good is said to be a **Giffen good**. Thus: $\left.\frac{\partial q_i}{\partial I}\right|_{dp=0} < 0$ and $\frac{dq_i}{dp_i} > 0$.

A Giffen good has a positively sloped Marshallian demand curve. In Figure 2.43 good 1 is Giffen:

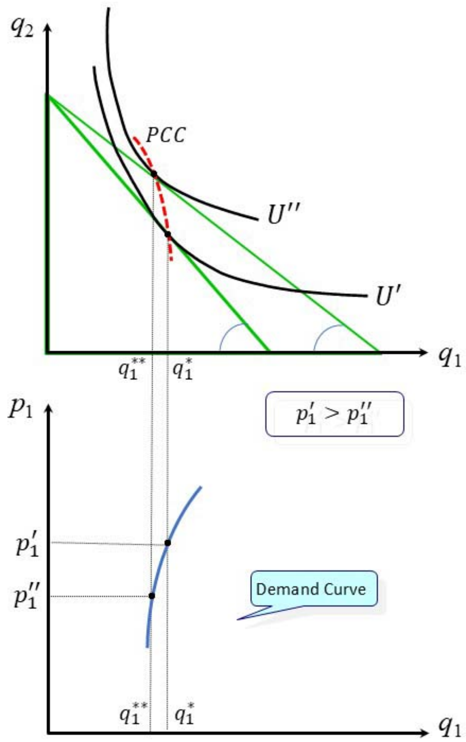

Figure 2.43 *Giffen good.*

2.17 Elasticities and the Slutsky equation

A problem with price and income effects is that they are not unit free: the fact that they depend on the units of measurement of each good makes comparisons across goods impossible. The notion of elasticity solves this problem, relating the ratio between rates of change to the ratio of levels. This results from ratios of log changes. That is:

$$\epsilon_{y,x} = \frac{dlny}{dlnx} = \frac{\frac{dy}{dx}}{\frac{y}{x}} = \frac{\frac{dy}{y}}{\frac{dx}{x}}$$

We define:

1. **Income elasticity** $\eta_i = \left(\frac{I}{q_i}\right)\left(\frac{\partial q_i}{\partial I_i}\right)$ where $\frac{\partial q_i}{\partial I_i}$ is the income effect;
2. **Own (uncompensated or Marshallian) price elasticity of demand** $\epsilon^M_{ii} = \left(\frac{p_i}{q_i}\right)\left(\frac{dq_i}{dp_i}\right)$ where $\frac{dq_i}{dp_i}$ is the total price effect. This elasticity is negative (except for Giffen goods), although we usually consider it in absolute value $\left|\epsilon^M_{ii}\right|$;
3. **Own (compensated or Hicksian) price elasticity of demand** $\epsilon^H_{ii} = \frac{p_i}{q_i}\left(\frac{\partial q_i}{\partial p_i}\right)$ where $\left.\frac{dq_i}{dp_i}\right|_{dU=0}$ is the substitution effect. This elasticity is negative, although we usually consider it in absolute value $\left|\epsilon^H_{ii}\right|$;
4. **Cross (uncompensated or Marshallian) price elasticity of demand** $\epsilon^M_{ij} = \left(\frac{p_j}{q_i}\right)\left(\frac{dq_i}{dp_j}\right)$;
5. **Cross (compensated or Hicksian) price elasticity of demand** $\epsilon^H_{ij} = \frac{p_j}{q_i}\left(\frac{\partial q_i}{\partial p_j}\right)$.

We also define:

1. If $\eta_i = \left(\frac{I}{q_i}\right)\left(\frac{\partial q_i}{\partial I_i}\right) > 0$ the good is normal;
2. If $0 < \eta_i = \left(\frac{I}{q_i}\right)\left(\frac{\partial q_i}{\partial I_i}\right) < 1$ the good is necessary;
3. If $\eta_i = \left(\frac{I}{q_i}\right)\left(\frac{\partial q_i}{\partial I_i}\right) > 1$ the good is luxury;
4. If $\eta_i = \left(\frac{I}{q_i}\right)\left(\frac{\partial q_i}{\partial I_i}\right) < 0$ the good is inferior.

Moreover:

1. If $\epsilon^H_{ij} = \left(\frac{p_j}{q_i}\right)\left(\frac{dq_i}{dp_j}\right) > 0$ the two goods are substitutes;
2. If $\epsilon^H_{ij} = \left(\frac{p_j}{q_i}\right)\left(\frac{dq_i}{dp_j}\right) < 0$ the two goods are complements.

We can redefine the Slutsky equation using elasticities, which is sometimes convenient especially on empirical grounds. Take (40) and multiply by p_i/q_i:

$$\frac{p_i}{q_i}\left(\frac{dq_i}{dp_i}\right) = \frac{p_i}{q_i}\left(\frac{dq_i}{dp_i}\right)_{dU=0} - \frac{p_i q_i}{I}\frac{I}{q_i}\left(\frac{dq_i}{dI}\right)$$

so that:

$$\epsilon^M_{ii} = \epsilon^H_{ii} - \omega_i \eta_i \tag{42}$$

where $\omega_i = \frac{p_i x_i}{I}$ is the share of income spent of good i. More generally:

$$\frac{dq_i}{dp_j} = \frac{p_j}{q_i}\left(\frac{dq_i}{dp_j}\right)_{dU=0} - \frac{p_j q_j}{I}\frac{I}{q_j}\left(\frac{dq_i}{dI}\right) \implies \epsilon^M_{ij} = \epsilon^H_{ij} - \omega_j \eta_i \tag{43}$$

Consider now (42). If the good is normal, then $|\epsilon^M_{ii}| > |\epsilon^H_{ii}|$; if the good is inferior, then $\left|\epsilon^M_{ii}\right| < |\epsilon^H_{ii}|$. We can draw Marshallian and Hicksian demand curves in the same graph. If the good is normal the Marshallian curve is more elastic than the Hicksian one; vice versa if the good is inferior.

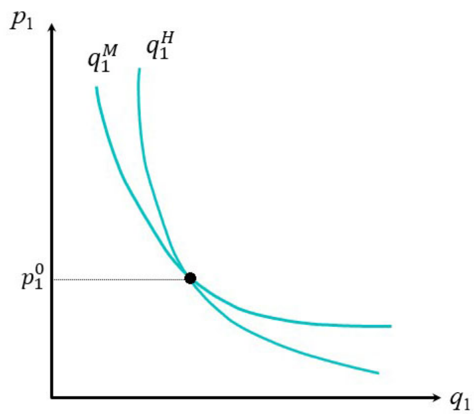

Figure 2.44 *Marshallian and Hicksian demand curves for normal goods.*

Take a reference price p_1^0 and and suppose the price increases. Corresponding to the new price we have $q_1^H - q_1^M > 0$. This implies that, in order to keep the consumer at the level of utility she had before the change, we must "compensate" her for the loss of purchasing power of her income. Hence the term "compensated" for Hicksian demand functions. Of course, for price reductions we would have to "uncompensate" or take some income away from the consumer. Finally, in the case of inferior goods the Hicksian demand curve is more price elastic than the Marshallian demand curve as the next figure shows.

Exercise: Compute the Marshallian demand functions under Cobb-Douglas preferences $U = q_1^a q_2^b$. Compute price and income elasticities for both goods. State any peculiarity exhibited by these preferences.

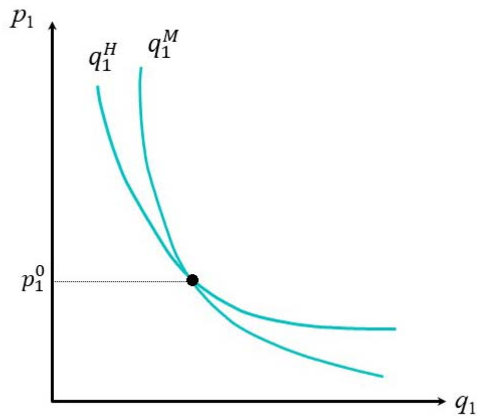

Figure 2.45 *Marshallian and Hicksian demand curves for inferior goods.*

2.18 The expenditure function

If we substitute the system of Hicksian demand functions (32) into the objective function (27) we obtain the optimal value function which is referred to as (minimized) **Expenditure Function** (EF). This function defines the minimum expenditure sufficient to attain a given level of utility by the consumer for given prices. The EF depends on the problem's exogenous variables. Thus:

$$E(p_1, p_2, U) = p_1 q_1 (p_1, p_2, U) + p_2 q_2 (p_1, p_2, U) \tag{44}$$

Chapter 3

Producer behavior: technology, cost and profit

3.1 The producer and the firm

In neoclassical production theory the producer is an agent who carries out a specific activity: transforming resources into other resources. Unlike modern companies the firm is simply an entity where this transformation takes place. The head and owner of the firm is the entrepreneur who is endowed with the ability to carry out such transformation using the available technology, buying inputs in appropriate markets and selling outputs in other markets. The reward to this managerial ability, sometimes called entrepreneurship, is a specific type of income: profits. In the circular flow model used to represent the structure and working of the economic system (see Figure 1.1), the entrepreneur is typically a member of the household (after all she is an individual who consumes, works and saves) obtaining profits (and possibly a salary if she works in the firm) as income.

3.2 Profit maximizing behavior

The basic behavioral assumption is that the firm/producer maximizes economic (not accounting) profits. Economic profits are the residual left after subtracting (all economic) production costs from revenues from sale. Costs arise from the purchase of labor services, capital equipment and structures, raw materials, energy and intermediate goods that are used in production. Economic costs differ from accounting costs because the latter typically do not include all costs of production factors. Revenues arise from sales of finished products or services or intermediate goods sold to other firms. The typical firm here will produce a manufactured (final or intermediate) good. We assume that there are no inventories, so that production and sales are the same thing.

Let a represent the vector of decision or choice variables available to the producer. Assuming they generally affect both revenues and costs, the problem is:

$$\max_{a} \Pi = R(a) - C(a) \tag{1}$$

where $a = (a_1, ..., a_n)$ are the endogenous or choice variables. Note that in order not to clutter the notation we will put no primes or "T superscripts" to denote vectors: we rely on the reader's ability to distinguish scalars and vectors given the context. The a vector will include input and output levels, input and output prices, and – in non-competitive contexts – advertising expenditures, output prices of competitors, capacity investment, and the like.

The general solution to (1) requires that a basic ("golden") rule be satisfied: **the marginal revenue and the marginal cost associated to each decision variable must be equal for all choice variables**. That is: $\partial R/\partial a_i = \partial C/\partial a_i$ or $MR_i = MC_i \; \forall i$. This is a necessary condition for optimality and holds for both perfectly competitive and imperfectly competitive input and output markets.

Remarks:

Remark 1. We can consider and study a few **special cases** of profit maximization in (1): (i) cost minimization (holding revenues constant); (ii) revenue maximization (holding costs constant); (iii) restricted profit maximization (holding some decision variable fixed or constrained at some given level);

Remark 2. Typically, both revenues and costs consist of two elements: **price and quantity**.

Profit maximizing decisions are subject to two types of **constraints**:

<u>*Market constraint*</u>: it pertains to the competitive nature of input and output markets where the firm operates, with special reference to whether or not the producer is able to affect the prices of inputs bought or outputs sold;

<u>*Technological constraint*</u>: it pertains to the feasibility of production plans and to the way the transformation activity is carried out. Specifically, we can represent it as follows:

$$\text{Inputs} \Longrightarrow \boxed{\text{Technology}} \Rightarrow \text{Outputs}$$

In terms of constraint 1 and of the remark 2 we will assume that the firm is **price-taker in all input and output markets**. A market where every household and every firm are price-takers is a **perfectly competitive market**. This assumption will be maintained throughout and in terms of (1) this involves choosing quantities of inputs and outputs for given prices. That is:

$$\max_{\{y,x\}} \Pi = p'y - w'x \tag{2}$$

where (p, w) are vectors of output and input prices and (y, x) are corresponding output and input quantities.

The study of producer behavior involves many analogies with consumer theory, especially in the case of expenditure/cost minimization. But there are also important differences. The description of the technology of the firm plays the same fundamental role as preferences for the consumer; however, the former is a constraint and the latter an objective. One important aspect to bear in mind is that, unlike the consumer, all variables are observed and measurable in producer theory.

Remark 3. Inputs and outputs are usually flow measures, i.e. measured with respect to a time interval. Often some inputs, e.g. fixed capital, are stock measures, measured at an instant in time. This is because data on utilization rates – such as machine time of use – which would translate stocks into flows of production services are unavailable. The implicit assumption for fixed capital is therefore that the utilization rate is equal to one, so that stock and flow are the same thing. More generally, inputs and outputs should be characterized by calendar time at which they are available, location, circumstances, and characteristics (quality). All these aspects are neglected here. All variables are taken to be continuous.

Remark 4. The analysis that follows is atemporal (static) so that all variables carry no time subscript. Although real firm decisions are inherently intertemporal, calling for a dynamic analysis of them, we maintain throughout the assumption that there is only a single time period – this was maintained also in the study of consumer behavior where no distinction between durable and nondurable consumption goods was made. We note that the technological constraint is such because at each instant in time there is a given set of production possibilities. Of course, these possibilities change over time owing to technological change causing the appearance (or disappearance) of new products and new production processes. Although neoclassical production theory has dealt with technological change, this aspect is beyond the scope of these notes.

3.3 Technological sets and the Production function

In order to study firm choices, we need a way to summarize the production possibilities that are available to the firm. The firm produces alternative levels and combinations of outputs using alternative levels and combinations of inputs. We aim to describe and characterize these different combinations. We begin with a definition.

3.3.1 Definitions of production sets

Production plan: a technologically feasible combination of input and output levels. At each point in time the state of technology determines and restricts the possibilities to combine inputs to produce outputs, and there are several ways by which we can represent this constraint.

Suppose that the firm has L possible commodities to serve as inputs and/or outputs. If the firm uses y_j^i units of a good j as an input and produces y_j^o of the good as an output, then the net output of good j is given by $y_j = y_j^o - y_j^i$ (note that i stands for input and o for output). This is sometimes referred to as a **netput**, as commodity j – a brick, say – may be the output of a furnace but also the input of the construction firm. A production plan is simply a list of netputs of various commodities. It will be a vector $y \in R^L$ where $y_j < 0$ if the j-th good serves as a net input and positive if the good serves as a net output. An example is $y = (7, -2, -3)$ which means that the firm produces 7 units of commodity one using 2 units of commoditiy two and 3 units of commodity three. This leads to the:

Production Possibility Set (or Production set or Technology set) Y: the set of all technologically feasible production plans that will be denoted by $Y \subset R^L$. The set describes all the possibile ways to transform inputs into outputs that are permitted by the current state of technology. It thus gives a complete description of the technological possibilities available to the firm. We can represent it as in Figure 3.1.

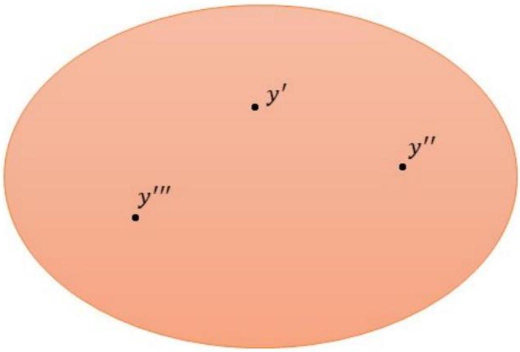

Figure 3.1 *Production possibilities set.*

We can illustrate the production set and production plans in various ways.

Example. In the case of only two commodities Figure 3.2 shows that all points below the curved line are feasible production plans. We have the following alternative situations: 1) commodity 1 can be used to produce commodity 2: $(y_1 < 0,\ y_2 > 0)$ (region A); 2) commodity 2 can be used to produce commodity 1 $(y_1 > 0,\ y_2 < 0)$ (regoin B); 3) nothing is done $(y_1 = 0,\ y_2 = 0)$ (origin 0); 4) both commodities can be used without producing an output$(y_1 < 0,\ y_2 < 0)$ (region C). Of course, the last situation is wasteful: no profit maximizing firm would ever choose to use inputs and incur costs without producing any output to get any revenue. Still, from a technical standpoint we allow for this possibility.

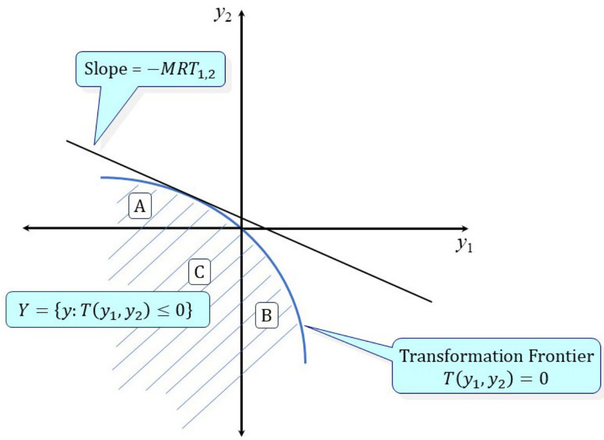

Figure 3.2 *Two-commodity production possibilities set.*

Restricted (or Short-run) Production Possibilities Set $Y(z)$. It hinges on the distinction between production plans that are "immediately" feasible and plans that are "eventually" feasible. There is a time interval, which we may call short-run, during which a subset of inputs $z \in R_+^p$ (with $p < n$; n is the number of inputs) is fixed at given levels, so that only production plans that are compatible with z are feasible in the short-run. All plans in Y become feasible in what we may call the long-run when all inputs including the z vector are variable so that their level can be freely changed. Thus $Y(z) \subset Y$ is a short-run production possibilities set. A production plan $(y, -k, -l)$ has output y produced using capital k and labor l as inputs. We can suppose that labor can be varied immediately but that capital is fixed (for some time, i.e. for the short-run) at level $\overline{k}$. Then: $Y\left(\overline{k}\right) = \left\{(y, -k, -l) \text{ in } Y : k = \overline{k}\right\}$ is the short-run production possibilities set.

Example 1. An example of restricted production possibilities set is where a firm uses a single input in fixed quantity to produce two outputs y_1 and y_2. This is shown in Figure 3.3.

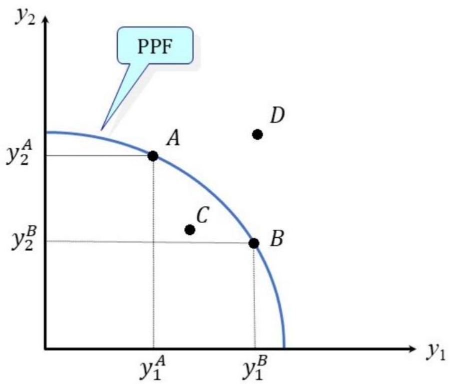

Figure 3.3 *One input-two outputs production possibilities set.*

Example 2. If we instead restrict the space to only one input and only one output (or we hold constant all inputs and outputs except one) we have the following representation of the production set.

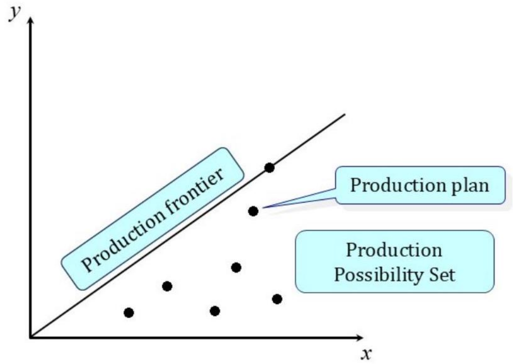

Figure 3.4 *One output-one input production possibilities set.*

In the case of single output technologies, it is customary to denote by x the input vector and y the single output.

Input Requirement Set $V(y)$: the set of all input bundles that produce at least a given bundle y of ouput units: $V(y) = \{x \in R^L : (y, -x) \in Y\}$. Note that the input requirement set as defined here measures inputs as positive numbers rather than negative numbers as used in the production possibilities set. Note also that $V(y)$ is an example of $Y(z)$ when z refers to outputs rather than inputs. Also, note that we may define a **Restricted or Short-run Input Requirement Set** $V(y, z)$.

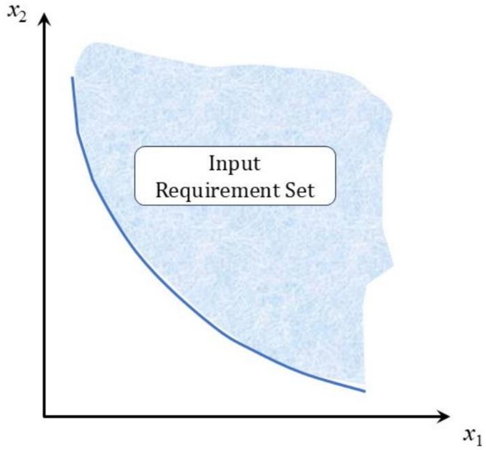

Figure 3.5 *Input requirement set.*

There is an alternative, more convenient way to represent the above production sets: it is by means of a **mathematical representation of production frontiers**.

Consider Figure 3.2. A rational profit maximizing firm will not make wasteful choices, given by those production plans that lie below the frontier and not on it. In fact, the firm could either produce more output using the same amount of input or the same output using less input. Either way, the firm would earn higher profit. We can call production plans that are below the frontier **inefficient production plans**. We say that a production plan $y \in Y$ is (technologically) efficient if there is no $y' \in Y$ such that $y' \geq y$ and $y' \neq y$: a production plan is efficient if there is no way to produce more output with the same inputs or the same output with less inputs. The set of efficient plans can be conveniently described with a mathematical function whose properties summarize the properties of the production set.

In Figure 3.2 the frontier is called **transformation function** $T \rightarrow R$ generally denoted with the implict function $T(y) = 0$. The transformation function is such that: 1) $T(y) = 0$ if y is on the frontier; 2) $T(y) < 0$ if y is in the interior of Y; 3) $T(y) > 0$ if y outside of Y. If $T(y) < 0$ y represents some sort of waste, although $T(.)$ tells us neither the form of the waste nor the magnitude. 4) the convexity of Y implies that $T(y)$ is convex or at least quasi-convex.

In Figure 3.3 the transformation function is typically referred to as **production possibilities frontier**, which in this case can be denoted as $T\left(y_1, y_2; \overline{x}\right) = 0$ where x is the single input and the bar refers to the fact that its quantity is given.

Remark: Transformation functions, which are appropriate for multi-output or multi-product technologies are represented as implicit functions, as there is no a priori reason to express one output as a function of other outputs as well as inputs. Doing so – for example writing $G(y) = H(x)$ – would amount to impose an unnecessary separability restriction on the technology.

In Figure 3.4 the frontier is commonly known as the **production function** $y = f(x_1)$ where y is the single output and x_1 is the single input. Note that as the production function picks out the maximum scalar output as a function of inputs, the transformation function picks out the maximal vector of net outputs. Of course, it is possible to represent the case of many inputs which in the figure are assumed to be set at a given level. Thus: $y = f(x)$ where x is a vector of inputs. Changes in the level of the inputs different from x_1 can be represented graphically with shifts of the production curve as Figure 3.6 shows.

If we stick to mathematical functions, we are not constrained by the dimensions (at most three) of graphical representations. In this case we define the production set as follows:

$$y = \left\{(y, -x) : y - f(x) \leq 0\right\}$$

where $y - f(x)$ is the boundary of set Y and where we include the equality sign because Y is assumed to be a closed set and so it includes his boundary (see below). The **production function** is the maximum feasible output given the input vector $y = f(x_1, ..., x_n)$.

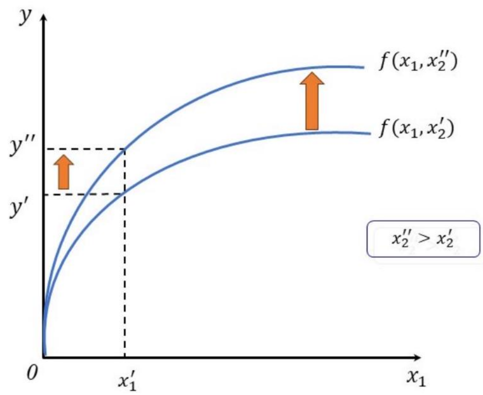

Figure 3.6 *Production curve.*

Finally, the frontier of the input requirement set in Figure 3.5 is known as the **isoquant**. Formally we define the isoquant as the locus of minimum input vectors used to produce a given output vector, that is:

$$Q(y) = \{x \text{ in } R_+^n : \; x \text{ is in } V(y) \text{ but not in } V(y') \; for \; \forall \; y' > y\}$$

where y is given. There is an isoquant for every output level, so that – like in the case of indifference curves – we have a map of isoquants.

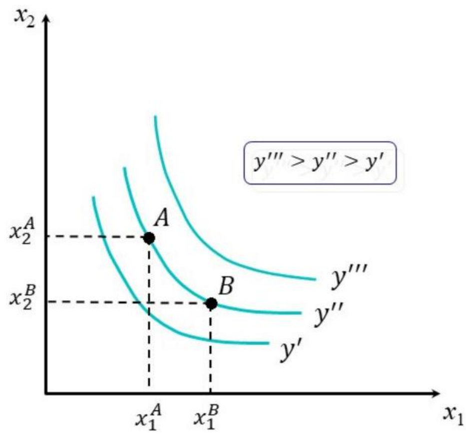

Figure 3.7 *Map of isoquants.*

In addition, isoquants can take on different shapes: see Figure 3.8.

A technology can be fully described as in the following examples for a single output y and two inputs (x_1, x_2).

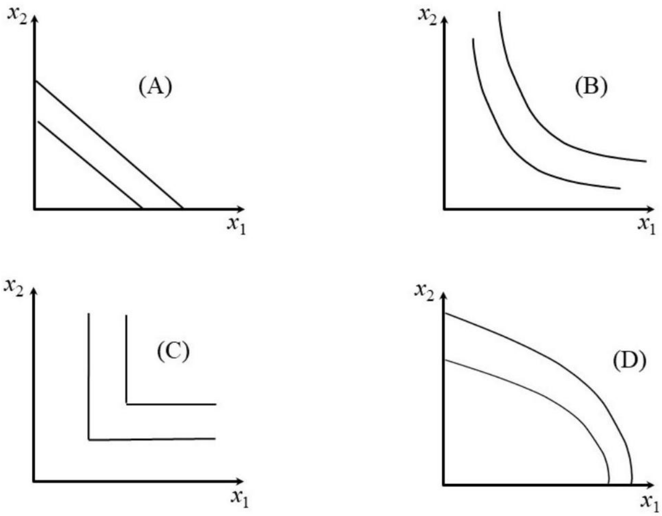

Figure 3.8 *Different shapes of isoquants.*

Example 1: **Leontief or Fixed coefficients technology**. Let $a > 0$ and $b > 0$ be parameters. Then:

$$Y = \{(y, -x_1, -x_2) \text{ in } R^3 : y \leq \min(ax_1, bx_2)\}$$

$$V(y) = \{(x_1, x_2) \text{ in } R_+^2 : y \leq \min(ax_1, bx_2)\}$$

$$Q(y) = \{(x_1, x_2) \text{ in } R_+^2 : y = \min(ax_1, bx_2)\}$$

$$T(y, x_1, x_2) = y - \min(ax_1, bx_2)$$

$$f(x_1, x_2) = \min(ax_1, bx_2)$$

Example 2: **Cobb-Douglas technology**. Let $a > 0$ and $b > 0$ be parameters. Then:

$$Y = \left\{(y, -x_1, -x_2) \text{ in } R^3 : y \leq x_1^a x_2^b\right\}$$

$$V(y) = \left\{(x_1, x_2) \text{ in } R_+^2 : y \leq x_1^a x_2^b\right\}$$

$$Q(y) = \left\{(x_1, x_2) \text{ in } R_+^2 : y = x_1^a x_2^b\right\}$$

$$T(y, x_1, x_2) = y - x_1^a x_2^b$$

$$f(x_1, x_2) = x_1^a x_2^b$$

The map of isoquants of the previous two examples is represented in the first two panels of Figure 3.8. The Cobb-Douglas parametrization has become extremely popular for its smooth representation and for its convenience as a didactic tool.

The Cobb–Douglas form was developed and tested against statistical evidence by Charles Cobb and Paul Douglas between 1927 and 1947. American politician and economist Paul Douglas explained that his first formulation of the Cobb–Douglas production function was developed in 1927; when seeking a functional form to relate estimates he had calculated for workers and capital, he spoke with mathematician and colleague Charles Cobb, who suggested the function $AL^{\beta}K^{1-\beta}$, previously used by Knut Wicksell, Philip Wicksteed, and Léon Walras.

3.3.2 Properties of production sets

To make progress we need to impose some structure on the above technology sets. This amounts to make a few fundamental assumptions that are general enough to be reasonable and valid for any production process. These assumptions play in producer theory the same role as axioms on preferences play in consumer theory.

A.1 **Regularity**. It entails the following:

a. Y is **non-empty** $\Rightarrow$ there is same way to produce any level of output. If Y is empty, then we have nothing to talk about;

b. $0 \in Y \Rightarrow$ **possibility of inaction**, i.e. no action on output is a possible production plan or the firm can choose to do nothing, in which case it would earn zero profits. Situations where $0 \notin Y$ arise when the firm has a fixed factor of production. For example, if the firm is obligated to pay rent on its factory, then it cannot do nothing. The cost of an unavoidable fixed factor of production is sometimes called a sunk cost. Note, however, that whether a cost item is fixed or not depends on the relevant time frame: if the firm waits long enough, its lease will expire and it will no longer have to pay its rent. Thus, while inaction is not a possibility in the short-run, it is a possibility in the long-run, provided that the long-run is sufficiently long.

c. No **free lunch**, i.e. we cannot produce output without using any input. In other words, any feasible production plan y must have at least one negative component. Beside violating the laws of physics, if there were a free lunch, then the firm could make infinite profits just by replicating the free lunch point over and over, which makes the firm's profit maximization problem impossible to solve.

d. Y is **closed** $\Rightarrow$ whenever a sequence of production plans $y^j \in Y$ $j = 1, 2, \dots$ and $y^j \to y^*$, then this limit production plan y^* is also in Y. This guarantees that the boundary of the set belongs to the set. Three implications: (i) closedness of Y implies that $V(y)$ is closed for all $y \geq 0$: if you have a sequence of input bundles x^i and each can produce y and $x^i \to x^*$, then x^* can produce y too; (ii) the production and transformation functions defined above are part of set Y; (iii) if a set does not contain its boundary, then if you try to maximize a function such as profit subject to the constraint that the production plan be in Y, there may be no optimal plan: the firm will try to be as close to the boundary as possible, but no matter how close it is, it could always be a little closer.

Remark: Regularity guarantees that (in the case of a single output) the **production function is continuous**. Same for the transformation function.

A.2 **Monotonicity** or **Free disposal**. If $y \in Y$ implies that $y' \in Y$ for all $y' \leq y$, then the set y is said to satisfy the free disposal or monotonicity property. This implies that commodities (either inputs or outputs) can be thrown away. This property means that Y includes all vectors where there are only inputs, but no outputs. The idea is that you can throw away as much as you want, and while you have to buy the commodities you are throwing away, you don't have to pay anybody to dispose of it for you. So, if there are two commodities, grapes and wine, and you can make 10 cases of wine from 1 ton of grapes, then it is also feasible for you to make 10 cases of wine from 2 tons of grapes (by just throwing one of ton of grapes away) or 5 cases of wine from 1 ton of grapes (by just throwing 5 cases of wine away at the end), or 5 cases of wine from 2 tons of grapes (by throwing 1 ton of grapes and 5 cases of wine away at the end).

A weaker requirement is that we only assume that the input requirement set is monotonic: if x is in $V(y)$ and $x' > x$ then x' is in $V(y) \Rightarrow$ you can certainly produce that output with more inputs.

Remark: (In the case of a single output) if the production function is differentiable, monotonicity ensures a **non-negative marginal productivity** for each input.

A.3 **Irreversibility**: $Y \bigcap \{-Y\} = \{0\}$ or: if $y \in Y$ and $y \neq 0$, then $-y \notin Y$. A production plan is not reversible unless it is a non-action plan. The laws of physics imply that all production processes are irreversible. You may be able to turn gold bars into jewelry and then jewelry back into gold bars, but in either case you use energy. So, this process is not really reversible.

A.4 **Additivity**: if y' and y'' are in Y, then $y^r = (y' + y'')$ is also in Y.

A.5 **Divisibility:** If y is in Y, then $y' = \alpha y$ is also in Y for $\alpha > 0$.

Assumptions A.4 and A.5 imply:

A.6 **Convexity:** If y and y' are in Y, then, for $0 \leq \alpha \leq 1$, $y^s = \alpha y + (1-\alpha) y'$ is in Y. This means that if x and x' can both produce y, then any weighted average of them can also produce y. Convexity of Y means that, if all goods are divisible, it is often reasonable to assume that two production plans y and y' can be scaled downward and combined. Convexity of the production set is a strong hypothesis if it holds everywhere. For example, convexity of the production set rules out the possibility of convex-concave technologies (increasing and then decreasing returns to scale: see later).

Remark 1. **Strict convexity**: if y and y' are in Y, then, for $0 < \alpha < 1$, $y^s = \alpha y + (1-\alpha) y$ is in $int\{Y\}$, the interior points of Y. Strict convexity guarantees that the profit maximizing production plan is unique provided it exists.

Remark 2: If the production set Y is a convex set, then the associated **input requirement set, V(y), is a convex set**. In fact, consider the production plans $s = (y, -x)$ and $s" = (y, -x")$ both in Y. Then by convexity of S: $s_{mz} = \alpha s + (1-\alpha)s" = \{\alpha y + (1-\alpha)y, -(\alpha x + (1-\alpha)x")\}$ is in Y. Hence the result follows.

Remark 3: **V(y) is a convex set** if and only if the production function $f(x)$ is a quasi-concave function. In fact, $V(y) = \{x : f(x) = y\}$ is just the upper contour set of $f(x)$. But a function is quasi-concave if and only if it has a convex upper contour set.

Remark 4: Convexity of the input requirement set **V(y)** ensures that isoquants are weakly convex. Strict convexity of **V(y)** implies strongly convex isoquants.

Note that some of the above properties – such as irreversibility – are not strictly necessary to get the results below. On the other hand, one should be aware of results that hinge on them.

3.3.3 Indicators of production sets

We now consider alternative ways to characterize a technology by means of its functional representation, the production function. In what follows we will be working exclusively with single output technologies. This is an assumption made for simplicity that enables us to fully characterize the firm's optimal decisions. A similar analysis adopting multi-output technologies is of course possible and it allows a richer set of results, but it is beyond the scope of these lecture notes. Occasionally we will refer to the case of multi-output technologies.

The technology indicators we will consider belong to two aspects: 1) the relationship between inputs, 2) the scale of production.

The starting point is a production function where one output y is produced using n inputs $x = (x_1, x_2, \ldots, x_n)$. We will often find it convenient to consider the two-input case for graphical representation. We begin by asking what is the change in the output level brought about by a change in all inputs. The answer is obtained by totally differentiating the production function $y = f(x_1, ..., x_n)$:

$$dy = \sum_{i=1}^{n} \omega_i dx_i \tag{3}$$

A marginal change in the level of production is given by a weighted average of the marginal changes in all input quantities. Consider now two special cases of (3).

Special case 1. Let the quantity of only one input change. Thus: $dx_j = 0 \ \ \forall j \neq i$. From (3) we get:

$$\omega_i = \frac{\partial f(x)}{\partial x_i} = MP_i(x) \tag{4}$$

Each "weight" ω_i represents the change in output brought about by a change in the quantity of an individual input, holding all remaining inputs constant. This notion is referred to as the **Marginal productivity** or **Marginal product of an input.** As noted above, marginal productivities are positive (non-negative) owing to axiom A.2 (monotonicity). We can therefore rewrite (3) as follows:

$$dy = \sum_{i=1}^{n} MP_i(x)dx_i \tag{5}$$

Note that marginal productivities are in general not constant but are functions of x themselves. For instance, if the production function is quasi-concave – due to the convexity property A.6 – having a convex portion first followed by a concave portion, then the marginal productivity of an input will be inverted-U shaped. This is represented in the two-good case in Figure 3.9.

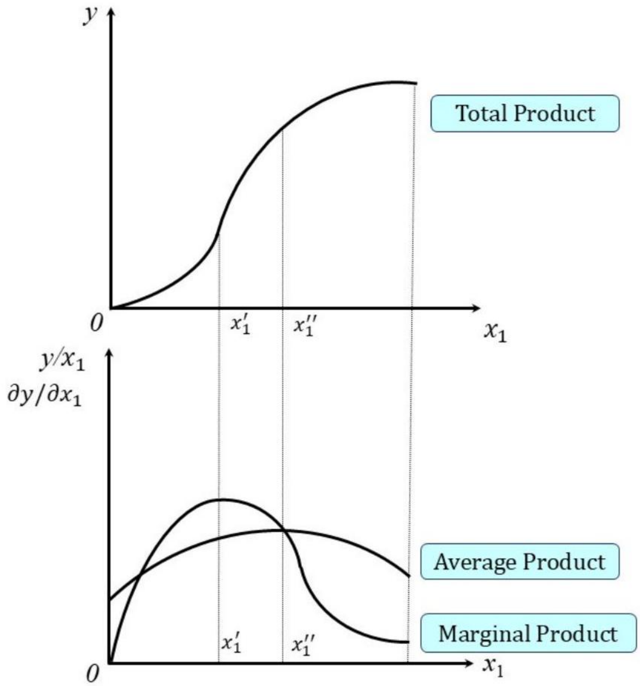

Figure 3.9 *Production function and Marginal productivity.*

For completeness, in Figure 3.9 we also report the **Average productivity of the input**, defined as $AP_i(x) = \frac{y}{x_i} = \frac{f(x)}{x_i}$, which is a function of x itself and bell-shaped. Indeed, recall that an average is the sum of (positive or negative) additions. When additions (the marginal elements) are positive and increasing the average goes up, when additions are positive but decreasing the average *after some time* starts going down (as the average has the memory of all past marginal additions).

In neoclassical growth models the production function must possess certain properties to ensure the stability of an economic growth path. These conditions, known as In **Inada conditions**, consist of the following assumptions about the shape of the production function:

- the value of the function at $x = 0$ is $f(0) = 0$;
- the limit of the first derivative (marginal productivity) when x approaches zero is $+\infty$: $\lim_{x\to 0} f_i(x) = +\infty$, meaning that the effect of the first unit of input has the largest effect;
- the limit of the first derivative (marginal productivity) when x approaches $+\infty$ is 0: $\lim_{x\to +\infty} f_i(x) = 0$, meaning that the effect of one additional unit of input is zero when approaching the use of infinite units of x_i;
- the production function is concave, i.e. the Hessian matrix is negative semi-definite, implying that the marginal productivities are positive but decreasing, i.e. $f_i > 0$ and $f_{ii} < 0$. Thus:

$$H = \begin{bmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{n1} & \cdots & f_{nn} \end{bmatrix}$$

is negative semi-definite. In the two-input case the determinant of H is: $|H| = f_{11}f_{22} - f_{12}^2 \geq 0$. Recall that convexity of production sets (Axiom A.6) only implies that the production function is quasi-concave, which is a milder curvature property relative to concavity. Concavity will be required when we tackle the firm's profit maximization problem, as we will see later, but it is not required for cost minimization. Quasi-concavity entails that the Bordered Hessian matrix B:

$$B = \begin{bmatrix} 0 & f_1 & \cdots & f_n \\ f_1 & f_{11} & & f_{1n} \\ \vdots & & \ddots & \vdots \\ f_n & f_{n1} & \cdots & f_{nn} \end{bmatrix}$$

is negative semi-definite. In the case of two inputs, the determinant of the Bordered Hessian B is given by: $|B| = -f_1^2 f_{22} + 2f_1 f_2 f_{12} - f_2^2 f_{11} \geq 0$.

Special case 2. Go back to the total differential (3) and allow only two inputs (x_i, x_j) to change. In addition, hold the level of output constant, so that $dy = dx_k = 0$ for $k \neq i, j$. We get: $f_i dx_i + f_j dx_j = 0$. Thus, the **Marginal rate of technical substitution** (MRTS):

$$MRTS_{ij}(x) = -\frac{dx_i}{dx_j} = \frac{f_j(x)}{f_i(x)} \tag{6}$$

is the ratio of the marginal products of the two inputs and measures the extent to which one input can be substituted with another one at unchanged level of output holding all remaining inputs constant. The property of isoquants being negatively sloped implies substitution between the two inputs in the two-input case. In addition, because isoquants have to be weakly convex, they can have linear portions or be uniformly linear.

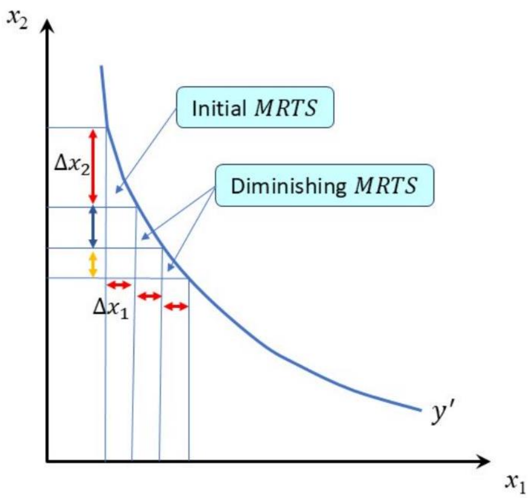

Figure 3.10 *Isoquants and the Marginal rate of technical substitution.*

As is apparent from Figure 3.10, the curvature of the isoquant represents the changing value of the MRTS, i.e. the degree of substitutability between inputs. Given its convexity, the MRTS varies along the isoquant and is diminishing in absolute value as we move down the curve. To show the convexity of an isoquant for two inputs we can look at how its slope changes, i.e. we can compute $dMRTS_{12}/dx_1 = d(f_1/f_2)/dx_1$:

$$d(f_1/f_2) = \frac{\left(f_{11}dx_1 + f_{12}dx_2\right)f_2 - \left(f_{21}dx_1 + f_{22}dx_2\right)f_1}{f_2^2}$$

Divide both sides by dx_1:

$$\frac{d\left(f_1/f_2\right)}{dx_1} = \frac{\left[f_{11} + f_{12}(dx_2/dx_1)\right]f_2 - \left[f_{21} + f_{22}(dx_2/dx_1)\right]f_1}{f_2^2}$$

Noting that $f_{12} = f_{21}$ and $dx_2/dx_1 = -f_1/f_2$:

$$\frac{d\left(f_1/f_2\right)}{dx_1} = \frac{f_2 f_{11} - 2f_1 f_{12} + (f_1^2/f_2)f_{22}}{f_2^2}$$

or:

$$\frac{d\left(f_1/f_2\right)}{dx_1} = \frac{1}{f_2^3}\left[f_2^2 f_{11} - 2f_1 f_2 f_{12} + f_1^2 f_{22}\right] \tag{7}$$

The term in square brackets is the determinant of the Bordered hessian which in turn accounts for the quasi-concavity of the production function. Hence the curvature of the production function implies a certain curvature of the isoquants.

There are a few special cases of isoquants worth mentioning that are portrayed in Figure 3.11.

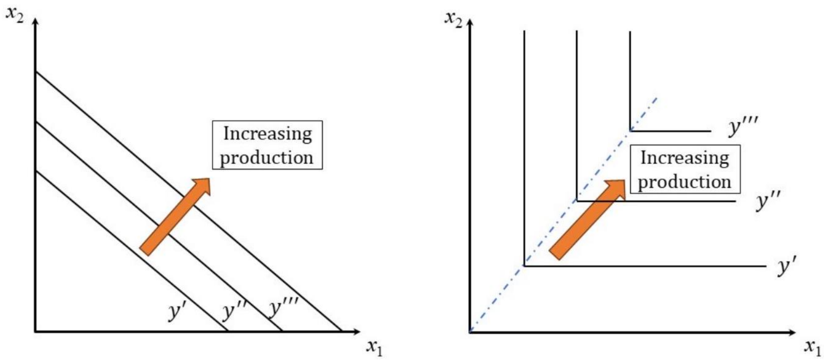

Figure 3.11 *Shape of isoquants and the degree of substitutability of inputs.*

On the left panel we have linear isoquants with a constant MRTS, implying infinite substitutability of inputs: that is the case of perfect substitutes. The corresponding production function is linear, so that in the two-input case: $y = a_1x_1 + a_2x_2$ and the $MRTS = -a_2/a_1$. On the right panel we have L-shaped isoquants, corresponding to fixed coefficients or Leontief technologies, which imply that only one combination of inputs will be used, the one at the kink. Here substitutability is zero and the MRTS along the isoquant is not uniquely defined. The production function in the two-input case is described in Example 1 above after Figure 3.8, that is: $y = \min(ax_1, bx_2)$.

Example. **MRTS for a Cobb-Douglas technology**. Given $y = Ax_1^a x_2^b$ (note that A is the level of output when a single unit of each input is applied; it is usually taken to be non zero and is taken to represent the "level" of the technology when looking at technological change), we can compute the marginal products. Then:

$$f_1(x_1, x_2) = aAx_1^{a-1}x_2^b = a\frac{Ax_1^a x_2^b}{x_1} = a\frac{y}{x_1}$$

$$f_2(x_1, x_2) = bAx_1^a x_2^{b-1} = b\frac{Ax_1^a x_2^b}{x_2} = b\frac{y}{x_2}$$

So that:

$$MRTS_{12} = \frac{a}{b}\frac{x_2}{x_1}$$

Note that the MRTS is proportional to the input mix in this case.

Special case 3. The problem with the MRTS as a measure of substitutability is that it is not informative when different inputs have different units of measurement so that it cannot be used to compare different technologies. The natural solution is to convert it into an elasticity measurement. This is what the **Elasticity of input substitution** σ does. It measures the percentage change in the input mix (or relative use of inputs) in response to a percentage change of the relative marginal products (or relative input prices), with output kept constant. Let $y = f(x_1, x_2)$, then:

$$\sigma = -\frac{dlog(x_1/x_2)}{dlog(f_1/f_2)} = -\frac{f_1/f_2}{x_1/x_2}\frac{d(x_1/x_2)}{d(f_1/f_2)} = -\frac{f_1 f_2(f_1 x_1 + f_2 x_2)}{x_1 x_2(f_2^2 f_{11} - 2f_1 f_2 f_{12} + f_1^2 f_{22})} \tag{8}$$

If the production function is quasi-concave, or equivalently isoquants are convex, the denominator of (8) is negative and the elasticity of substitution is a positive number. To see how formula (8) is obtained rewrite the first expression as follows:

$$\sigma = -\left(\frac{f_1 x_2}{f_2 x_1}\right)\frac{d(x_1/x_2)/dx_1}{d(f_1/f_2)/dx_1}$$

Consider the second term on the right-hand side. The denominator $d(f_1/f_2)/dx_1$ is given by (7), whereas the numerator is given by:

$$\frac{d(x_1/x_2)}{dx_1} = \frac{1}{x_2^2}\left(x_2 - x_1\frac{dx_2}{dx_1}\right) = \frac{1}{x_2^2}\left(x_2 + x_1\frac{f_1}{f_2}\right) = \frac{1}{f_2 x_2^2}\left(f_1 x_1 + f_2 x_2\right)$$

Putting everything together:

$$\sigma = -\left(\frac{f_1 x_2}{f_2 x_1}\right)\frac{f_1 x_1 + f_2 x_2}{f_2 x_2^2}\frac{f_2^3}{f_2^2 f_{11} - 2f_1 f_2 f_{12} + f_1^2 f_{22}} = -\frac{f_1 f_2(f_1 x_1 + f_2 x_2)}{x_1 x_2(f_2^2 f_{11} - 2f_1 f_2 f_{12} + f_1^2 f_{22})}$$

Note that, since $f_{12} = f_{21}$, we have $\sigma_{12} = \sigma_{21}$, that is the elasticity is symmetric. In addition, note that it has nothing to do with prices, it is just a technological concept, although under cost minimizing behavior $\sigma = -\frac{dln(x_1/x_2)}{dln(f_1/f_2)} = -\frac{dln(x_1/x_2)}{dln(w_1/w_2)}$ since optimality requires $MRTS_{12} = w_1/w_2$ (see below). From (8) we see that the elasticity of substitution is related to the curvature of isoquants. Graphically, σ relates the (log) change in the $MRTS_{12} = f_1/f_2$ to the (log) change in the input mix x_1/x_2. Figure 3.12 illustrates the concept.

Extreme values of σ in the two-input case are seen when either the denominator or the numerator or the term $\frac{d(x_1/x_2)/dx_1}{d(f_1/f_2)/dx_1}$ in (8) is equal to zero:

a) $\sigma = 0$: Leontief or fixed coefficient technology, as the input mix does not change: $d(x_1/x_2) = 0$;

b) $\sigma = \infty$: linear technology, as the $MRTS_{12}$ is constant and therefore does not change: $d(f_1/f_2) = 0$.

The value σ is therefore a measure of the substitutability of inputs**:** $\sigma = 0$ entails no substitution, whereas $\sigma = \infty$ entails infinite substitutability.

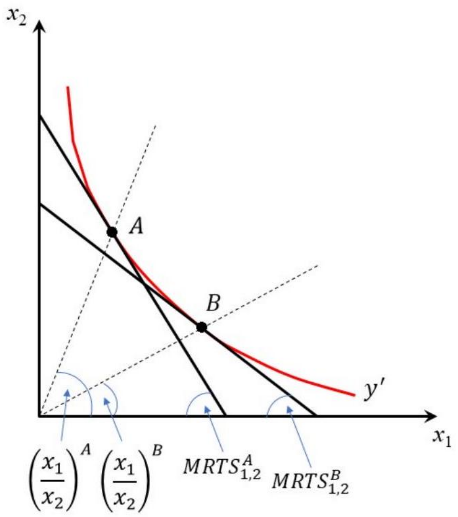

Figure 3.12 *Elasticity of input substitution.*

Example 1: **Elasticity of substitution for a Cobb-Douglas technology**. We have seen above that:

$$MRTS_{12} = \frac{a}{b}\frac{x_2}{x_1}$$

or:

$$\frac{x_2}{x_1} = \frac{b}{a} MRTS_{12}$$

Taking logs:

$$\ln\left(\frac{x_2}{x_1}\right) = ln\left(\frac{b}{a}\right) + lnMRTS_{12}$$

which implies:

$$\sigma_{12} = \frac{dln(x_2/x_1)}{dlnMRTS_{12}} = 1$$

The Cobb-Douglas technology entails a unitary elasticity of substitution among inputs, for all pairs of inputs. This is an unwarranted restriction which has prompted economists to formulate less restrictive production functions. A famous one is the so-called Constant Elasticity of Substitution (CES) production function introduced in 1956 by Solow for the two-factor (capital, labor) case and made popular in 1961 by Arrow, Chenery, Minhas, and Solow himself.

Example. **Constant Elasticity of Substitution (CES) technology**. Let a be a distribution parameter, $\rho > 0$ a substitution parameter, and ν a scale parameter. Then:

$$y = \left[a x_1^{\rho} + (1-a) x_2^{\rho}\right]^{\nu/\rho}$$

$$MRTS_{12} = \frac{a}{1-a}\left(\frac{x_1}{x_2}\right)^{\rho-1}$$

Taking logs:

$$\ln\left(\frac{x_2}{x_1}\right) = \frac{1}{1-\rho}\ln\left(\frac{a}{1-a}\right) + \frac{1}{1-\rho}\ln MRTS_{12}$$

which implies:

$$\sigma_{12} = \frac{dln(x_2/x_1)}{dlnMRTS_{12}} = \frac{1}{1-\rho}$$

Quite interestingly, if $\rho \to 0$ then the CES tends to the Cobb–Douglas production function; if $\rho \to \infty$ then the CES tends to the Leontief production function (inputs are perfect complements); $\rho \to 1$ then the CES tends to the linear production function (inputs are perfect substitutes).

It can be seen that σ is no longer constrained to be unity, although in the CES case the elasticity is still a constant value and the same for any input pair. This makes the CES still quite restrictive from a production theory standpoint. Further developments have led to the introduction of the so-called **Flexible Functional Forms** of which notable examples are the **Transcendental Logarithmic or Translog** form, the **Generalized Leontief**, and the **Quadratic** forms (and several others). These are beyond the scope of the present notes.

When we move to a world of more than two inputs expression (8) carries over to the more general case but becomes a measure of substitutability of a pair of inputs, holding all the other input levels constant. Thus $\sigma_{ij} = -\frac{dlog(x_i/x_j)}{dlog(f_i/f_j)}$ ($\forall\ dx_k = 0\ k \neq i, j$) is referred to as the **Allen-Uzawa partial elasticity of input substitution** (AUES). Note that we now have $-\infty \leq \sigma_{ij} \leq +\infty$, so that two inputs may be either substitutes of complements. Unfortunately, in the general n-input case there is more than one notion of elasticity of substitution. In particular, we can define a σ_{ij} with constant level of the other factors and a σ_{ij} with constant level of factor prices (Hicks and Morishima elasticities of substitution).

The third technology indicator we consider deals with the **Scale of production**. Recall the total differential of the production function (3). It expressed the marginal change in the level of production brought about by a weighted average of the marginal changes in all input quantities. We now specialize this relationship by considering proportional changes in all inputs tx_i for $t \geq 0$ and look at how output proportionally changes, say sy for $s \geq 0$. More specifically we want to see $s \lesseqgtr t$. This leads to the concept of **Returns to scale** of the technology. Thus:

1. **Constant returns to scale** (CRS) when $s = t$ or $tf(x) = f(tx)\ \forall t \geq 0$. In this case doubling (halving, tripling, etc.) all inputs doubles (halves, triples, etc.) output;
2. **Increasing returns to scale** (IRS) when $s > t$ or $tf(x) > f(tx)\ \forall t > 0$. In this case doubling (halving, tripling, etc.) all inputs more than doubles (halves, triples, etc.) output;
3. **Decreasing returns to scale** (DRS) when $s < t$ or $tf(x) < f(tx)\ \forall t > 0$. In this case doubling (halving, tripling, etc.) all inputs less than doubles (halves, triples, etc.) output.

The production set Y exhibits increasing returns to scale if any feasible production plan $y \in Y$ can be scaled down (up): $ay \in Y$ for $a \geq 1$. This means that a technology that exhibits increasing returns to scale is one that becomes more productive (on average) as the size of the output grows. The production set Y exhibits decreasing returns to scale if any feasible production plan $y \in Y$ can be scaled down: $ay \in Y$ for $a \in [0,1]$. This means that a technology that exhibits decreasing returns to scale is one that becomes less productive (on average) as the size of the output grows. Finally, Y exhibits constant returns to scale if it exhibits both decreasing returns to scale and increasing returns to scale at all y. That is, for all $a \geq 0$, if $y \in Y$, then $ay \in Y$. Constant returns to scale means that the firm's productivity is independent of the level of production. Thus, it means that any feasible production plan can either be scaled upward or downward.

There is an important implication of returns to scale for the firm's cost structure. Indeed, a firm that is characterized by (local) increasing/constant/decreasing returns to scale exhibits (local) decreasing/constant/increasing average or unit costs of production.

Note that the various kinds of returns to scale defined above are global in nature. In fact, most real technologies exhibit none of these. The "typical" technology is one that at first exhibits increasing returns to scale and then exhibits decreasing returns to scale. This would be the case, for example, for a manufacturing firm whose factory size is fixed. At first, as output increases, its average productivity increases as it spreads the factory cost over more output. However, eventually the firm's output becomes larger than the factory is designed for. At this point, the firm's average productivity falls as the workers become crowded, machines become overworked, etc. A typical example of this type of technology is illustrated in the upper panel of Figure 3.9.

When returns to scale vary depending on the level of output a local measure of returns to scale is useful. This leads to the notion of **Elasticity of scale**. The elasticity of scale measures the percent increase in output due to one percent increase in all inputs due to an increase in the scale of operations. Given $y = f(x)$ being the production function, let t be a positive scalar and consider the function $y(t) = f(tx)$. If $t = 1$ we have the current scale of operation; if $t > 1$, we are scaling all inputs up by t; and if $t < 1$, we are scaling all inputs down by t. The elasticity of scale is given by:

$$e(x) = \left.\frac{dy(t)/y(t)}{dt/t}\right|_{t=1}$$

Rearranging this expression yields:

$$e(x) = \left.\frac{dy(t)}{dt}\frac{t}{y}\right|_{t=1} = \left.\frac{df(tx)}{dt}\frac{t}{f(tx)}\right|_{t=1} \tag{9}$$

Note that we must evaluate the expression at $t = 1$ to calculate the elasticity of scale at the point x. Thus, we have the following:

(Local) Returns To Scale. A production function $f(x)$ is said to exhibits locally increasing, constant, or decreasing returns to scale as $e(x) \gtreqless 1$.

3.3.4 Homogeneous production functions

There is an important class of production functions which are characterized by global returns to scale. They are the mathematical class of homogenous functions. Homogenous functions occur in many cases in consumer and producer theories, especially in duality theory as we will see.

Math digression: Homogenous functions
A continuous differentiable function $y = f(x)$ is said to be homogenous of degree k, i.e. it is O^k, if:

$$f(tx) = t^k f(x)$$

for $t > 0$. If $k = 1$ the function is said to be linear homogenous. Homogenous functions enjoy several properties, the most important for our purposes being the following ones:

1. Transmission through differentiation. If $f(x)$ is O^k, then $f_i(x)$ is O^{k-1}, $f_{ij}(x)$ is O^{k-2}, and so on. In fact, given $f(tx) = t^k f(x)$, we compute: $\frac{\partial f(tx)}{\partial x_i}\frac{\partial tx_i}{\partial x_i} = t^k \frac{\partial f(x)}{\partial x_i}$ which implies: $t\frac{\partial f(tx)}{\partial x_i} = t^k \frac{\partial f(x)}{\partial x_i}$ and in turn: $\frac{\partial f(tx)}{\partial x_i} = t^{k-1}\frac{\partial f(x)}{\partial x_i}$.

2. Scale invariance of ratios of partial derivatives: $\frac{f_i(tx_i,tx_j)}{f_j(tx_i,tx_j)} = \frac{t^{k-1} f_i(x_i,x_j)}{t^{k-1} f_j(x_i,x_j)} = \frac{f_i(x_i,x_j)}{f_j(x_i,x_j)}$.

3. Euler's theorem: $kf(x) = \sum_{i=1}^n f_i(x)x_i$. In fact, differentiate the homogenous function with respect to t: $kt^{k-1}f(x) = \sum_{i=1}^n \frac{\partial f(tx)}{\partial tx_i}\frac{\partial tx_i}{\partial t}$. Evaluate at $t = 1$. Note that, given property (1), Euler's theorem applies also to partial derivatives of the homogenous function. Thus: $(k-1)f_i(x) = \sum_{j=1}^n f_{ij}(x)x_j \,\forall i$.

Homogenous production functions have global degrees of returns to scale identified by the degree of homogeneity. Therefore, a production function $y = f(x)$ that is homogenous of degree k is characterized by (global) constant returns to scale (CRS) if $k = 1$, increasing returns to scale (IRS) if $k > 1$, and decreasing returns to scale (DRS) if $k < 1$.

Linear homogeneous technologies represent an interesting special case. A technology enjoys (global) **constant returns to scale** only if it is linear homogeneous. CRS is also equivalent to state that $y \in Y$ implies $ty \in Y \quad \forall t \geq 0$ or, equivalently,

$x \in V(y)$ implies $tx \in V(ty)\ \forall t > 1$. Another result is worth noting. Owing to Euler's theorem, in the two-input case: $y = f(x) = f_1 x_1 + f_2 x_2$ and in turn:

$$\begin{cases} 0 = f_{11}x_1 + f_{12}x_2 \\ 0 = f_{12}x_1 + f_{22}x_2 \end{cases} \implies \begin{array}{l} f_{11} = -\frac{f_{12}x_2}{x_1} \\ f_{22} = -\frac{f_{12}x_1}{x_2} \end{array} \tag{10}$$

Consider the denominator of (8) and use (10):

$$\begin{aligned} x_1 x_2 \left(f_2^2 f_{11} - 2 f_1 f_2 f_{21} + f_1^2 f_{22}\right) = & \\ &= -x_1 x_2 \left(f_2^2 \frac{f_{12}x_2}{x_1} + 2 f_1 f_2 f_{21} + f_1^2 \frac{f_{12}x_1}{x_2}\right) = \\ &= -f_{12}\left(f_2^2 x_2^2 + 2 f_1 f_2 x_1 x_2 + f_1^2 x_1^2\right) = \\ &= -f_{12}\left(f_1 x_1 + f_2 x_2\right)^2 = \\ &= -f_{12} y^2 \end{aligned}$$

Thus:

$$\sigma = -\frac{f_1 f_2 y}{-f_{12} y^2} = \frac{f_1 f_2}{y f_{12}} \tag{11}$$

which is the elasticity of input substitution for a linear homogeneous technology.

Examples. Let's see if the following production functions are homogenous, and if so, which is the degree of returns to scale characterizing them.

- $y = A x_1^a x_2^b \implies A(tx_1)^a (tx_2)^b = A t^{a+b} x_1^a x_2^b = t^{a+b} y$ homogenous of degree $a+b$;
- $y = \left[a x_1^\rho + (1-a) x_2^\rho\right]^{v/\rho} \implies \left[a(tx_1)^\rho + (1-a)(tx_2)^\rho\right]^{v/\rho} = \left\{t^\rho \left[a x_1^\rho + (1-a) x_2^\rho\right]\right\}^{v/\rho} = t^v y$ homogenous of degree v;
- $y = ax_1 + bx_2 \implies atx_1 + btx_2 = ty$ homogenous of degree 1;
- $y = ax_1 + bx_2 + cx_1 x_2 \implies atx_1 + btx_2 + ct^2 x_1 x_2$ non-homogenous.

3.3.5 Separability

When the set of inputs is large a natural question is whether they can be grouped. Technologies use different energy inputs (oil, gas, electricity, sun, wind, etc.), labor inputs (skilled, unskilled; blue and white collars, etc.), fixed capital inputs (equipment, structures, etc.), raw materials, intermediate inputs. Moreover, if an input is difficult to measure, we may wonder whether we can leave it aside if the focus of the analysis is on other aspects of the technology. Can a production process be separated in different stages so that at each stage individual inputs can be used to form an intermediate aggregate input, which is turn used with other aggregate inputs at a higher stage, and so on? If the answer to these questions is positive, we say that the technology is (functionally) separable.

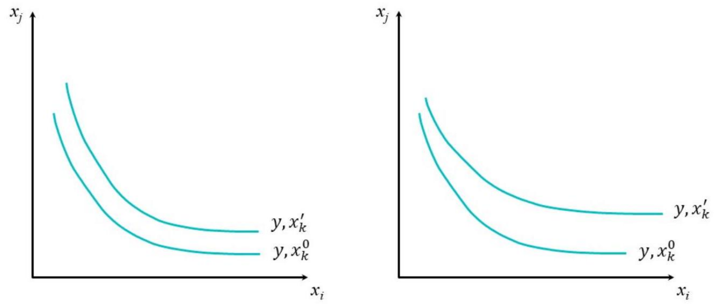

Figure 3.13 *Separability.*

If x_i and x_j are separable from x_k in $f(x)$, then $\frac{\partial}{x_k} MRTS_{ij} = \frac{\partial}{x_k} \frac{f_i(x)}{f_j(x)} = 0$. Graphically, this means that the slope of the isoquant in the (x_i, x_j) space is not affected by x_k: see Figure 3.13. Note that $\frac{\partial}{x_k} \frac{f_i(x)}{f_j(x)} = 0$ implies that $\frac{f_{ik} x_k}{f_i} = \frac{f_{jk} x_k}{f_j}$ or $\frac{\partial ln f_i}{\partial ln x_k} = \frac{\partial ln f_j}{\partial ln x_k}$, i.e. the elasticity of the marginal product of input x_i with respect to x_k is equal to the elasticity of the marginal product of input x_j with respect to x_k.
Partition the x vector into subsets indexed by $\{I^1, I^2, .., I^N\}$. We have:

1. **Weak separability**: $\frac{\partial}{x_k} \frac{f_i(x)}{f_j(x)} = 0 \quad I, j \in I^r, k \notin I^r$. In this case $MRTS_{ij}$ is independent of all inputs belonging to I^r. The production function can be written as: $f(x) = F\left[g^1\left(x^1\right), g^2\left(x^2\right), .., , g^N\left(x^N\right)\right]$.

2. **Strong separability**: $\frac{\partial}{x_k} \frac{f_i(x)}{f_j(x)} = 0 \qquad i \in I^r, i \in I^v,\ k \notin I^r \cup I^v$. In this case $MRTS_{ij}$ is independent of all inputs belonging to I^r and I^v when the two inputs belong to different subsets. The production function can be written as: $f(x) = F\left[\sum_{k=1}^{N} g^k\left(x^k\right)\right]$.

3.3.6 Multi-output technologies and the Marginal rate of transformation

In Section 3.1 we introduced the **transformation function** which is appropriate for multi-output technologies. Also the transformation function $T(y) = 0$ can be used to investigate how various inputs or outputs can be substituted for each other in the production process. We can define the **Marginal rate of transformation** (MRT) of commodity j for commodity i at y:

$$MRT_{ij} = \frac{\partial y_i}{\partial y_j} = -\frac{T_i(y)}{T_j(y)} \tag{12}$$

This is the slope of the transformation frontier with respect to the two commodities and it tells how much one must increase the (net) usage of factor j if one decreases the net usage of factor i in order to remain on the transformation frontier. It is important to note that factor usage can be either positive or negative in this model, representing either an output or an input of the production process. Thus (12) can be a **Marginal rate of technical substitution** if i, j are both inputs, a **Marginal rate of product transformation** if i, j are both outputs, and a **Marginal rate of intensity use** if i is output and j is input. Elasticities of substitution, transformation, and intensity can be correspondingly defined.

3.4 Profit maximization

As stated in Section 3.2, the firm's objective is to maximize profit. Recalling that in a multi-output context y represents both inputs and outputs, we may state the general profit maximization problem as follows:

$$\max_y \Pi = p \cdot y \quad \text{subject to} \quad T(y) = 0$$

Forming the Lagrangian $L = py - \xi T(y)$, the FOCs are:

$$\begin{cases} p_i = \xi T_i\left(y^*\right) \ \forall i \\ T(y^*) = 0 \end{cases}$$

where the asterisk denotes evaluation at the optimal point. Note that the first FOC implies:

$$\frac{p_i}{T_i\left(y^*\right)} = \xi = \frac{p_j}{T_j\left(y^*\right)} \quad \forall i, j$$

or:

$$\frac{p_i}{p_j} = \frac{T_i\left(y^*\right)}{T_j\left(y^*\right)} \quad \forall i, j \tag{13}$$

Optimality requires that the marginal rate of transformation – in the case of outputs – or the marginal rate of substitution – in the case of inputs – be equal to the corresponding price ratio. We illustrate the situation with a graph drawn for two netputs (one input, one output).

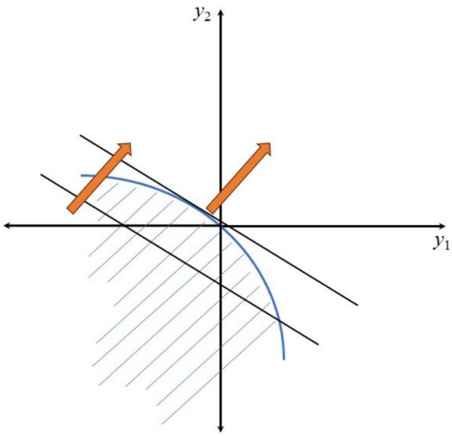

Figure 3.14 *Profit maximization.*

The curve represents the transformation function and the shaded area is the production possibilities set. The two lines are **isoprofit lines** or combinations of inputs and outputs for which the producer obtains a given level of profits. Note that $\Pi^0 = p_1 y_1 - p_2 y_2$ or $y_2 = \Pi^0 / p_2 - (p_1 / p_2) y_1$, so that the slope of the isoprofit line is the ratio of prices. Profits are maximized when the highest isoprofit line is tangent to the transformation function. This occurs where the marginal rate of transformation between the two netputs is equal to their price ratio. Hence, we move the profit line up and to the right until we find the point of tangency. This is the point that satisfies the optimality condition (13). Note that unless Y is convex, the first order conditions will not be sufficient for a maximum. Generally, second order conditions will need to be checked. This can be seen in Figure 3.14 where there is only a point of tangency.

The solution to the profit maximization problem, $y = y(p)$, is called the firm's **net supply function** and the value function for the profit maximization problem, $\Pi(p) = p \cdot y(p)$, is called the **profit function**.

One thing to worry about is whether this solution exists. Problems can arise when the firm's production set is not convex, i.e., its technology exhibits nondecreasing returns to scale. In this case, the production frontier is convex rather than concave as in Figure 3.14. In this case, if the firm maximizes profit, it either produces nothing at all (if the output price is small relative to the input price), or else it can always increase profit by moving further out on its production frontier. Thus, its production becomes infinite, as do its profits. As a result, we will focus most of our attention on technologies that exhibit decreasing returns.

In the familiar case of <u>single output</u> technologies, the profit maximization problem is the following:

$$\max_{\{y,x\}} \Pi = py - w'x \quad subject\ to \quad y = f(x) \tag{14}$$

The analog of the marginal rate of transformation is the marginal rate of technical substitution and the FOCs analogous to (13) are:

$$p_i f_i\left(x^*\right) = w_i \quad \forall i \tag{15}$$

This implies:

$$\frac{f_i\left(x^*\right)}{f_j\left(x^*\right)} = \frac{w_i}{w_j} \quad \forall i, j \tag{16}$$

From (16) we see that profit maximization requires the equality between the MRTS and the input price ratio for every input pair. The solution of (15) yields the optimal input demand functions. Plugging back these values into the objective function yields the maximum profit function and into the production function yields the optimal output supply function. We fully develop this analysis in section 3.8 below for the case of two inputs.

In general, we prefer to decompose the optimization problem into two stages. In the first stage we hold the level of output constant, set equal to an exogenously given level y^0, in which case the profit maximization problem $\max_{\{x\}} \Pi = py^0 - w'x \sim \min_{\{x\}} C = w'x$ conditional on y^0. This is the cost minimization problem. In the second stage we relax the constraint on output and choose the level that maximizes profit. Mathematically, it should be recalled that an optimization problem with many choice variables $z = \{z_1, z_2, .., z_n\}$ can always be solved one-shot with respect to all z_i $(i = 1, \dots, n)$ or it can be solved in subsequent stages, by optimizing first with respect to z_1 holding $\{z_2, .., z_n\}$ constant and forming the first-stage optimized objective; this represents the objective which is then optimized with respect to z_2 holding $\{z_3, .., z_n\}$ constant, and so on. At the end of this process the optimal value function is the same as the one obtained with the one-shot solution.

3.5 Cost minimization and Conditional input demand functions

We study the problem of choosing the input quantities which minimize the cost of producing a given level of output y^0. The solution of this problem will yield the cost-minimizing input demand functions and the (minimum) cost function. Before starting four remarks are in order.

1. This problem is completely analogous to the consumer expenditure minimization problem, just replace x with q, y with U, and C with E, although there are a few interesting specific results here.

2. This problem can be completely characterized also in the case of multi-output technologies, under the constraint of a given output vector. The minimum cost function $C(w, y)$ to be obtained below will depend on the vector of input prices and the vector of output quantities.

3. While profit maximization always implies cost minimization, the opposite is not necessarily true. Indeed, the firm may minimize the cost of producing a level of output which is not the profit-maximizing one.

4. There are practical cases where the firm is indeed constrained by the amount of output it has to produce. The main example is local public transport where the firm producing transportation services – bus, subway, etc. – is obliged to serve all its demand. In this case the best the firm can do is to minimize the cost of supplying the given amount of service.

In the two-input case the problem is to $\min_{\{x_1, x_2\}} C = w_1 x_1 + w_2 x_2$ subject to $f(x_1, x_2) = y^0$. We form the Lagrangian function:

$$\min_{\{x_1, x_2, \mu\}} \mathscr{L} = w_1 x_1 + w_2 x_2 + \lambda \left[y^0 - f(x_1, x_2) \right] \tag{17}$$

The FOCs are:

$$\begin{cases} \mathscr{L}_1 = w_1 - \lambda f_1(x_1^*, x_2^*) = 0 \\ \mathscr{L}_2 = w_2 - \lambda f_2(x_1^*, x_2^*) = 0 \\ \mathscr{L}_\lambda = y^0 - f(x_1^*, x_2^*) = 0 \end{cases} \tag{18}$$

where asterisks denote the optimal level and λ the Lagrange multiplier. Take the ratio of the first two equations to get:

$$MRTS(x_1^*, x_2^*) = \frac{f_1(x_1^*, x_2^*)}{f_2(x_1^*, x_2^*)} = \frac{w_1}{w_2} \tag{19}$$

$$y^0 = f(x_1^*, x_2^*) \tag{20}$$

According to the FOCs the optimal combination of inputs must be such that their marginal rate of technical substitution is equal to the ratio of their prices. Note that (19) is the same as (16). This shows that profit maximization always entails cost minimization. Graphically, (18) implies the tangency between the given isoquant and the lowest isocost line. Moreover, the optimal input combination must lie on the given isoquant. This is shown in Figure 3.15.

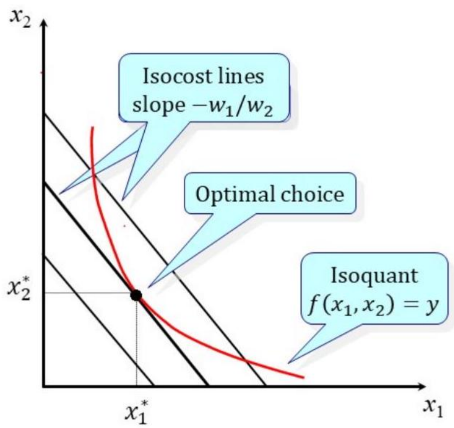

Figure 3.15 *Cost minimization.*

Note that the optimal input mix (x_1^*/x_2^*) can be seen as the angle of the line joining the origin and the point of optimal choice. The SOCs of the problem require the matrix:

$$A = \begin{bmatrix} -\lambda f_{11} & -\lambda f_{12} & -f_1 \\ -\lambda f_{12} & -\lambda f_{22} & -f_2 \\ -f_1 & -f_2 & 0 \end{bmatrix}$$

to be negative semi-definite. It is worth computing the determinant of A:

$$|A| = \lambda(f_1^2 f_{22} - 2 f_1 f_2 f_{22} + f_2^2 f_{11}) \leq 0 \tag{21}$$

This expression is nothing but the Bordered Hessian of the production function and is necessarily negative (non-negative) due to its (weak) quasi-concavity: see also (8). Thus, the sufficient condition for cost minimization requires the production function to be quasi-concave (or the input requirement set to be convex) in turn implying convex isoquants.

Remark. The SOC for cost minimization suggests two alternative strategies: 1) we may impose no curvature condition on the technology at the outset to later discover that for the minimization of costs quasi-concavity is required; 2) we may set up the cost minimization problem with a quasi-concave production function constraint implying that there is no need to check the second order conditions, as they are automatically satisfied.

We now restrict (21) to be strictly negative – or the production function to be strictly q-concave – so that we are guaranteed that the IFT holds and the FOCs system (18) can be solved for the optimal quantities:

$$\begin{cases} x_1 = \phi^1(w_1, w_2, y^0) \\ x_2 = \phi^2(w_1, w_2, y^0) \\ \lambda = \phi^\lambda(w_1, w_2, y^0) \end{cases} \tag{22}$$

The first two equations are referred to as **Conditional input demand functions**. They relate the optimal quantities of the two inputs to prices and output level. These demands are "conditional" on the given output level and are structurally analogous to consumer Hicksian demand functions.

The third equation represents the change in the objective function. As previously noted, Lagrange multipliers in these optimization problems always carry an economic interpretation. Indeed, λ has the important interpretation of **Marginal cost function** (MC). To prove this, it is enough to invoke the Envelope theorem as λ represents the change in the objective C due to a change in a parameter or exogenous variable y^0. As an exercise, let us review the proof. The envelope theorem states that, given:

$$\min f(x;\alpha) - \lambda(y - g(x;\alpha))$$

we have:

$$\frac{\partial f^*}{\partial \alpha} = \frac{\partial f}{\partial \alpha} + \lambda \frac{\partial h}{\partial \alpha} \quad \text{where} \quad h = y - g(x;\alpha)$$

Now let $C = w_1 x_1 + w_2 x_2$ and C^* be the minimized cost function. Thus:

$$\min C(.) + \lambda\left[(y - f(.))\right] \quad \Rightarrow \quad \frac{\partial C^*}{\partial y} = \frac{\partial C}{\partial y} + \lambda \frac{\partial\left[(y - f(.))\right]}{\partial y} = 0 + \lambda * 1 = \lambda$$

Thus:

$$\frac{\partial C^*}{\partial y} = MC = \lambda = \phi^\lambda(w_1, w_2, y^0) \tag{23}$$

3.6 Comparative statics of cost minimization, normal and regressive, substitute and complementary inputs

Totally differentiating the FOCs:

$$\begin{cases} dw_1 - \lambda f_{11} dx_1 - \lambda f_{12} dx_2 - f_1 d\lambda = 0 \\ dw_2 - \lambda f_{21} dx_1 - \lambda f_{22} dx_2 - f_2 d\lambda = 0 \\ dy - f_1 dx_1 - f_2 dx_2 = 0 \end{cases}$$

or in matrix form:

$$\begin{bmatrix} -\lambda f_{11} & -\lambda f_{12} & -f_1 \\ -\lambda f_{21} & -\lambda f_{22} & -f_2 \\ -f_1 & -f_2 & 0 \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \\ d\lambda \end{bmatrix} = \begin{bmatrix} -dw_1 \\ -dw_2 \\ -dy \end{bmatrix} \tag{24}$$

Let $|A|$ be the determinant of the matrix which by the SOCs must be a positive definite matrix, i.e. $|A| < 0$. By Cramer's rule we have:

Own price effect. Let $dw_1 \neq 0$ and $dw_2 = dy^0 = 0$ and solve for $\frac{dx_1}{dw_1}$:

$$\frac{dx_1}{dw_1} = \frac{\begin{vmatrix} -1 & -\lambda f_{11} & -f_1 \\ 0 & -\lambda f_{22} & -f_2 \\ 0 & -f_2 & 0 \end{vmatrix}}{|A|} = \frac{f_2^2}{|A|} < 0 \tag{25}$$

$$\frac{dx_2}{dw_2} = \frac{f_1^2}{|A|} < 0 \tag{26}$$

Conditional input demand functions are unambiguously negatively related to their own price. Graphically, the situation is the following:

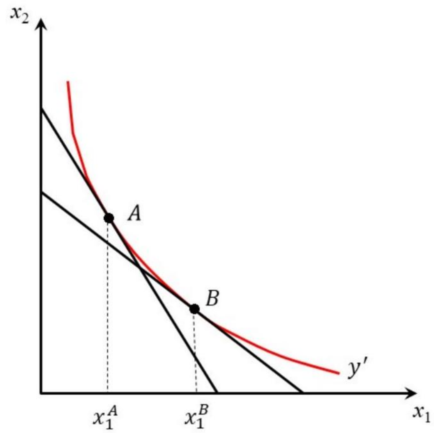

Figure 3.16 *Change in input price and optimal input combination.*

We see that the price change leads the isocost line to change slope and the equilibrium point to move along the given isoquant. The isocost line becomes less steep if the price of input 1 decreases, thereby inducing an increase in the quantity demanded. It is important to note that the negative slope of isoquants is responsible for the negative price effect which portrays a **substitution effect** between inputs.

Cross price effect. Let $dw_2 \neq 0$ and $dw_1 = dy^0 = 0$ and solve for $\frac{dx_1}{dw_2}$:

$$\frac{dx_1}{dw_2} = \frac{\begin{vmatrix} 0 & -\lambda f_{11} & -f_1 \\ -1 & -\lambda f_{22} & -f_2 \\ 0 & -f_2 & 0 \end{vmatrix}}{|A|} = -\frac{f_1 f_2}{|A|} > 0 \tag{27}$$

$$\frac{dx_2}{dw_1} = -\frac{f_1 f_2}{|A|} > 0 \tag{28}$$

This effect is unambiguously positive, meaning that when the price of input 2 goes up, and its demand goes down, the demand for input 1 goes up. There are two remarks to be made here.

Cross price effects of conditional input demand functions are symmetric: $\frac{dx_i}{dw_j} = \frac{dx_j}{dw_i}$. In empirical work this property can either be imposed by suitable parametrizations of input demand functions or, alternatively, represent a testable implication of the theory.

In the case of only two inputs, these can only be substitutes. More generally, for every pair of n inputs:

- If the cross price effect is positive $\frac{dx_i}{dw_j} > 0$ the two goods are **substitutes**;
- If the cross price effect is negative $\frac{dx_i}{dw_j} < 0$ the two goods are **complements**;

With more than two goods, holding the quantity of all other goods constant, we may have cross price effects for any pair of goods that are either positive or negative (or equal to zero).

It is often convenient to work with price elasticities. The own price elasticity of input demand is given by $\epsilon_{ii} = \frac{\partial x_i}{\partial w_i}\frac{w_i}{x_i}$ and is always negative, whereas the cross price elasticity of input demand is $\epsilon_{ij} = \frac{\partial x_i}{\partial w_j}\frac{w_j}{x_i} \lesseqgtr 0$. Note that cross price elasticities are no longer symmetric.

We now consider the:

Output effect. Let $dy \neq 0$ and $dw_1 = dw_2 = 0$ and solve for $\frac{dx_1}{dy}$:

$$\frac{\partial x_1}{\partial y} = \frac{\begin{vmatrix} 0 & -\lambda f_{11} & -f_1 \\ 0 & -\lambda f_{22} & -f_2 \\ -1 & -f_2 & 0 \end{vmatrix}}{|A|} = -\frac{\lambda(f_{12} f_2 - f_{22} f_1)}{|A|} \lesseqgtr 0 \tag{29}$$

$$\frac{\partial x_2}{\partial y} = -\frac{\lambda(f_{11} f_2 - f_{12} f_1)}{|A|} \lesseqgtr 0 \tag{30}$$

These effects tell us which factor is used more or less than the other one as output increases. In general, they should be positive but this is not a rule. Indeed, in the case of n inputs:

- If the output effect is positive $\frac{dx_i}{dy} > 0$ the input is said to be **normal**;
- If the output effect is negative $\frac{dx_i}{dy} < 0$ the input is said to be **regressive** (or **inferior**).

A little more than a curiosity is the fact that in the case of just two inputs they cannot be both inferior inputs. To show this totally differentiate the production function $dy = f_1 dx_1 + f_2 dx_2$ and divide throughout by dy. This yields $1 = f_1 \frac{dx_1}{dy} + f_2 \frac{dx_2}{dy}$ which shows that $\frac{dx_i}{dy}$ $i = 1,2$ cannot be both negative.

We may also define the output elasticity of input demand functions as follows: $\varepsilon_{iy} = \frac{\partial x_i}{\partial y} \frac{y}{x_i} \lesseqgtr 0$.

We now focus on the comparative statics for the **marginal cost of output**. Firstly, we look at the effects that changes in input prices have on marginal cost. We note that:

$$\frac{\partial \lambda}{\partial w_1} = -\frac{\lambda(f_{12} f_2 - f_{22} f_1)}{|A|} \lesseqgtr 0 \tag{31}$$

$$\frac{\partial \lambda}{\partial w_2} = -\frac{\lambda(f_{12} f_1 - f_{11} f_2)}{|A|} \lesseqgtr 0 \tag{32}$$

Note that the numerators of (29) and (31) are the same; similarly for (30) and (32). Indeed, if the price of an input increases, marginal cost increases. This implies that the input is normal. Vice versa for regressive inputs.

Turning to the impact of a change in output on marginal cost we have:

$$\frac{\partial \lambda}{\partial y} = -\frac{\lambda^2(f_{11} f_{22} - f_{12}^2)}{|A|} \lesseqgtr 0 \tag{33}$$

It is interesting to note that producing more may increase, decrease or leave marginal cost unchanged.

Graphically:

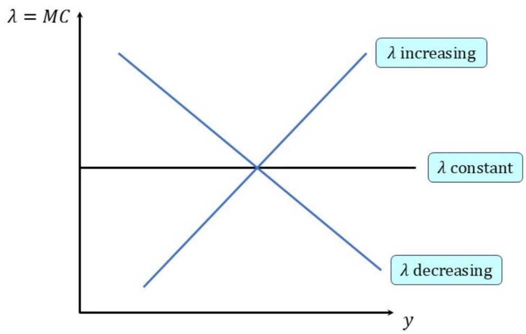

Figure 3.17 *Impact of output change on marginal cost.*

Hence, our theory does not constraint the behavior of marginal cost as production levels change. In other words, the marginal cost curve may slope upward, downward or be horizontal.

It is of interest to ask under which conditions $MC = \lambda$ is independent of the level of output, i.e. $\frac{\partial \lambda}{\partial y} = 0$ (holding factor prices constant). Consider (33). Since λ is positive and the determinant $|A|$ negative, it must be the case that: $f_{11}f_{22} - f_{12}^2 = 0$.

Consider a linear homogeneous production function. Expression (10) implies:

$$\begin{cases} 0 = f_{11}x_1 + f_{12}x_2 \\ 0 = f_{12}x_1 + f_{22}x_2 \end{cases} \implies \begin{bmatrix} f_{11} & f_{12} \\ f_{12} & f_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{34}$$

Or, in matrix form, $A \cdot x = 0$. Because $x \neq 0$ we must have $|A| = f_{11}f_{22} - f_{12}^2 = 0$. We coinclude that a **linear homogeneous production function has constant marginal cost**. Graphically, for constant returns to scale homogenous production functions the marginal cost curve is horizontal.

3.7 The cost function

If we substitute the system of conditional input demand functions (22) into the objective function (17) we obtain the value function which is referred to as (minimized) **Cost Function** (CF). This function defines the minimum cost the producer must bear when producing a given level of output for given input prices. The cost function depends on the problem's exogenous variables. Thus:

$$C\left(w_1, w_2, y\right) = w_1\phi^1(w_1, w_2, y) + w_2\phi^2(w_1, w_2, y) \tag{35}$$

3.8 Profit maximization, unconditional input demand and output supply functions

So far we have minimized the cost of producing a given level of output. Nothing, however, guarantees that that level of output is the one that maximizes profits. To find the profit maximizing level of output we can set up a simple, unconstrained, single choice variable optimization problem. Using the minimized cost function (35) the problem is:

$$\max_y \Pi = py - C\left(w_1, w_2, y\right) \tag{36}$$

The FOC and the SOC are:

$$\frac{\partial \Pi}{\partial y} = p - \frac{\partial C\left(w_1, w_2, y^*\right)}{\partial y} = 0 \tag{37}$$

$$\frac{\partial^2 \Pi}{\partial y^2} = -\frac{\partial^2 C\left(w_1, w_2, y^*\right)}{\partial y^2} \leq 0 \tag{38}$$

The optimal level of output is found where the (exogenous) output price equals marginal cost. This is to occur on the upward sloping portion of the marginal cost curve, as the SOC requires the marginal cost function to be convex in output at the optimal output level, i.e. $\frac{\partial^2 C}{\partial y^2} \geq 0$. Graphically:

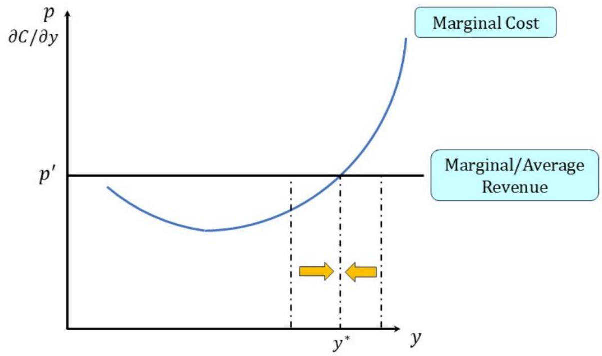

Figure 3.18 *Profit maximizing output level.*

If we restrict the SOC to be strictly negative, or the cost function to be strictly convex in output, then by the IFT we may solve the FOC for output to obtain:

$$y = y(w_1, w_2, p) \tag{39}$$

This is the firm's optimal **output supply function**. This can be plugged into the profit maximand (34) to obtain maximum profits $\Pi(w_1, w_2, p)$: see Section 3.10 below.

The problem with the profit maximization problem (36) is that it obscures the underlying choice of optimal input quantities. If we want to study this aspect, we may revert back to problem (14). Given the linear nature of the objective function, we may easily substitute the technological constraint into the objective to set up the unconstrained profit maximization problem:

$$\max_{\{x_1, x_2\}} \Pi = pf(x_1, x_2) - w_1 x_1 - w_2 x_2 \tag{40}$$

The FOCs are:

$$\begin{cases} \dfrac{\partial \pi}{\partial x_1} = pf_1(x_1^*, x_2^*) - w_1 = 0 \\ \dfrac{\partial \pi}{\partial x_2} = pf_2(x_1^*, x_2^*) - w_2 = 0 \end{cases} \tag{41}$$

Equilibrium requires that for each input the **Value of the marginal product** (the contribution to revenue from sale of additional output obtained from an increase in the input use) pf_i is equal to its cost (the market price) w_i. Alternatively, the physical marginal product has to be equal to the real cost of each factor: $f_i = w_i/p$. Recalling that marginal productivity was represented graphically in Figure 3.9, the profit maximizing choice of an input is shown in Figure 3.19.

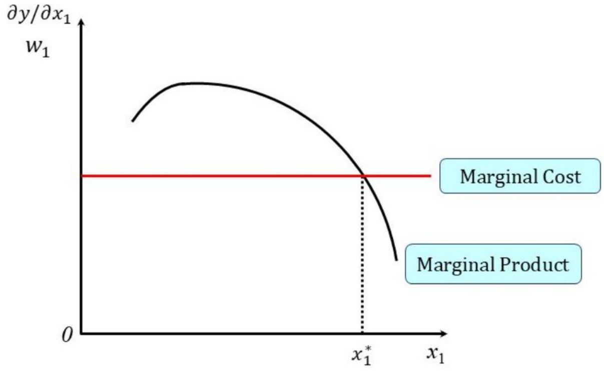

Figure 3.19 *Profit maximizing input level.*

Note that the ratio of the two FOCs (41) implies that the $MRTS_{12} = \frac{f_1}{f_2} = \frac{w_1}{w_2}$ and this holds for any pair of inputs. This is the optimality condition of cost minimization (19), thus proving that profit maximization implies cost minimization.

The SOCs are such that:

$$D = \begin{bmatrix} pf_{11} & pf_{12} \\ pf_{21} & pf_{22} \end{bmatrix}$$

is positive semi-definite. After noting that $pf_{11}, pf_{22} < 0$, the determinant of D should be:

$$|D| = p(f_{11}f_{22} - f_{12}^2) \geq 0 \tag{42}$$

It can be seen that $\left(f_{11}f_{22} - f_{12}^2\right)$ is the determinant of the Hessian of the production function, so that the SOCs of profit maximization require the production function to be concave. The curvature condition on the technology is thus stronger for profit maximization than for cost minimization (a concave function is also quasi-concave but the opposite is not true). Note also that the optimal input choice must be located in the declining portion of the marginal product curve as shown in Figure 3.18 and required by $pf_{ii} < 0$.

Strengthening (42) to positive definite – $f_{11}f_{22} - f_{12}^2 > 0$ – implies that the IFT holds and the FOC system (41) can be solved explicitly to produce the **Unconditional input demand functions:**

$$\begin{cases} x_1 = \varphi^1(w_1, w_2, p) \\ x_2 = \varphi^2(w_1, w_2, p) \end{cases} \tag{43}$$

As a further step we can substitute the optimal input demand functions (43) in the technological constraint to get:

$$y = f((x_1^*, x_2^*,) = f\left[\varphi^1\left(w_1, w_2, p\right), \varphi^2\left(w_1, w_2, p\right)\right] = \psi\left(w_1, w_2, p\right) \tag{44}$$

This is the firm's profit maximizing **Output supply function** (sometimes also referred to as **Indirect production function**).

3.9 Comparative statics of profit maximization

Totally differentiating the FOCs (41):

$$\begin{cases} pf_{11}dx_1 + pf_{12}dx_2 - dw_1 + f_1 dp = 0 \\ pf_{21}dx_1 + pf_{22}dx_2 - dw_2 + f_2 dp = 0 \end{cases}$$

Or:

$$\begin{bmatrix} pf_{11} & pf_{12} \\ pf_{21} & pf_{22} \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix} = \begin{bmatrix} dw_1 & -f_1 dp \\ dw_2 & -f_2 dp \end{bmatrix} \tag{45}$$

We study the impact of changes in input and output prices on the optimal input choice.

Own (input) price effect. Let $dw_1 \neq 0$ and $dw_2 = dp = 0$ and solve for $\frac{dx_1}{dw_1}$: Thus:

$$\frac{dx_1}{dw_1} = \frac{\begin{vmatrix} 1 & pf_{11} \\ 0 & pf_{22} \end{vmatrix}}{|D|} = \frac{pf_{22}}{|D|} < 0 \tag{46}$$

$$\frac{dx_2}{dw_2} = \frac{pf_{11}}{|D|} < 0 \tag{47}$$

These effects are unambiguously negative owing to the concavity of the production technology and imply that the unconditional input demand functions are inversely related to their own price. The demand curves are downward sloped.

Cross (input) price effect. Let $dw_2 \neq 0$ and $dw_1 = dp = 0$ and solve for $\frac{dx_1}{dw_2}$: Thus:

$$\frac{dx_1}{dw_2} = \frac{pf_{12}}{|D|} > 0 \tag{48}$$

$$\frac{dx_2}{dw_1} = \frac{pf_{12}}{|D|} > 0 \tag{49}$$

Note that also here the cross price effects are symmetric. Note also that in principle there is no a priori sign attached to the cross partial f_{12} though it is usually taken to be positive so that the two inputs are substitutes.

Output price effect. Let $dp \neq 0$ and $dw_1 = dw_2 = 0$ and solve for $\frac{dx_1}{dp}$: Thus:

$$\frac{\partial x_1}{\partial p} = \frac{p(f_2 f_{12} - f_1 f_{22})}{|D|} \gtrless 0 \tag{50}$$

$$\frac{\partial x_2}{\partial p} = \frac{p(f_1 f_{12} - f_2 f_{11})}{|D|} \lessgtr 0 \tag{51}$$

Somewhat unexpectedly the signs are ambiguous. However, recall from the comparative statics of cost minimization that:

$$\frac{\partial x_1}{\partial y} = -\frac{\lambda(f_{12}f_2 - f_{22}f_1)}{|A|} \lesseqgtr 0 \tag{29}$$

$$\frac{\partial x_2}{\partial y} = -\frac{\lambda\left(f_{11}f_2 - f_{12}f_1\right)}{|A|} \lesseqgtr 0 \tag{30}$$

Comparing (50)-(51) and (29)-(30) we get:

$$\begin{cases} \frac{\partial x_1}{\partial p} = -\frac{p}{\lambda}\frac{|A|}{|D|}\frac{\partial x_1}{\partial y} \lesseqgtr 0 \\ \frac{\partial x_2}{\partial p} = -\frac{p}{\lambda}\frac{|A|}{|D|}\frac{\partial x_2}{\partial y} \lesseqgtr 0 \end{cases} \tag{52}$$

where $-\frac{p}{\lambda} < 0$ and $\frac{|A|}{|D|} < 0$. Thus:

$$\text{sign}\left(\frac{\partial x_1}{\partial p}\right) = \text{sign}\left(\frac{\partial x_1}{\partial y}\right) \tag{53}$$

We conclude that:

- In the case of a normal input, i.e. $\frac{\partial x_i}{\partial y} > 0$, the outpt price effect is positive $\frac{\partial x_i}{\partial p} > 0$;
- In the case of a regressive input, i.e. $\frac{\partial x_i}{\partial y} < 0$, the outpt price effect is negative $\frac{\partial x_i}{\partial p} < 0$.

We may ask at this point how the output supply function (44) behaves when the output price p changes. Differentiating the with respect to p we get:

$$\frac{\partial y}{\partial p} = \frac{\partial y}{\partial p} = f_1\frac{\partial x_1^*}{\partial p} + f_2\frac{\partial x_2^*}{\partial p} \tag{54}$$

Using the comparative statics results (50)-(51) into (54) we obtain:

$$\begin{aligned} \frac{\partial y}{\partial p} &= p\left[\frac{f_1\left(f_2f_{12} - f_1f_{22}\right)}{|D|} + \frac{f_2\left(f_1f_{12} - f_2f_{11}\right)}{|D|}\right] = \\ &= -p\left[\frac{f_1^2f_{22} - 2f_1f_2f_{22} + f_2^2f_{11}}{|D|}\right] = -p\left[\frac{|A|}{|D|}\right] > 0 \end{aligned} \tag{55}$$

Recalling the SOCs (21) and (42) we conclude that unambiguously the output supply function varies directly with the output price, or the output supply curve slopes upward.

One theoretically interesting (practically inconsequential) case arises with linear homogenous CRS production functions. From (34) we know that $f_{11}f_{22} - f_{12}^2 = 0$. But from (42) this implies $|D| = 0$. The determinant of the SOCs is singular, in which case the IFT does not hold and we cannot solve explicitly the FOCs (41) for the optimal input demand functions. It follows that price-taking behavior, profit maximization and constant returns technologies lead to undefined unconditional factor demands and undefined output supply. In this situation profits are unbounded.

3.10 The profit function

We conclude by plugging the optimal input demand functions (43) into the profit maximand:

$$\begin{aligned}\Pi = & \, pf\left[(\varphi^1(w_1, w_2, p), \varphi^2(w_1, w_2, p))\right] - w_1\varphi^1(w_1, w_2, p) - w_2\varphi^2(w_1, w_2, p) = \\ & = \Pi(w_1, w_2, p)\end{aligned} \tag{56}$$

This is referred to as the profit function and defines the maximum profit obtainable by the produce for given input and output prices. It depends on the problem's exogenous variables.

Chapter 4

Consumer behavior: duality between utility and expenditure and between utility and indirect utility

4.1 Duality in consumer behavior

We have previously characterized the behavior of the consumer based on her preferences as represented by the utility function and her budget constraint producing optimal decision in terms of demand functions for commodities. In so doing we have also derived the consumer expenditure and indirect utility functions. In this chapter we ask whether we can characterize the behavior of the consumer directly by means of her expenditure or indirect utility function to the extent that they fully reflect the properties of the original utility function and therefore of her preferences.

What we purport to show is that (under certain regularity conditions) the expenditure function $E(p,\ U)$ completely describes the preferences of the consumer; i.e., the agent's expenditure (often also called cost) function $E(p,\ U)$ can be used in order to define the consumer's utility function $U(q)$. Thus, there is a duality between expenditure and utility functions in the sense that either of these functions can describe the preferences of the individual equally under certain circumstances. The same holds for the duality between utility and indirect utility functions $V(p, I)$.

The strategy we follow is to derive the regularity conditions that an indirect utility function must have and to show how a "direct" utility function can be constructed from a "well behaved" indirect utility function. We will then rely on the indirect utility function so validated to obtain the optimal decision rules, i.e. the

Marshallian demand functions, in a simple way. We will do the same for the expenditure function.

To lay out the problem assume that there are n goods that are exchanged on each market. Generally stated, the consumer's problem is:

$$\max_{\{q\}} U(q) \text{ s.t. } pq \leq I, q \geq 0 \tag{1}$$

We assume for simplicity that all (n-dimensional) quantities are strictly positive, so that the second constraint is not relevant. In addition, the first constraint holds as a strict equality, since the present analysis is single period and rational behavior leads the consumer to spend all her income. The (six) axioms on preferences guarantee that equilibrium q^* is unique and varies continuously with p and I.

The conventional approach (the "primal" problem) to solve the consumer optimization problem entails the following steps:

1. Calculation of $n+1$ First order conditions (FOCs);

2. Check that the Second order conditions (SOCs) are satisfied, requiring to compute a $(n+1)\cdot(n+1)$ determinant: this step is not necessary if the convexity axiom is invoked at the outset, entailing the quasi-concavity of the utility function $U(q)$;

3. Inversion of the FOCs to obtain the complete system of Marshallian demand functions: this requires the strict convexity axiom to hold or the strict quasi-concavity of $U(q)$ as the implicit function theorem holds in this case (alternatively check and make sure that the Jacobian of the FOCs is non-singular);

4. Total differentiation of the FOCs to study the comparative statics of the problem and calculate via Cramer's rule the effect on dq of changes in dp and dI. This requires calculating ratios of $(n+1)\cdot(n+1)$ determinants;

5. Substitution of the Marshallian demand functions into the problem's objective function to obtain the maximized indirect utility function $V(p,I)$. In mathematical optimization problems these are usually referred to as **optimal value functions**.

It is clear that this process is extremely tedious and burdensome. In addition, it may well be the case that for certain parametrizations of the utility function to analytically obtain the demand functions from the FOCs (step 3) is not possible.

Ideally, we would like to have an approach that takes us to the agent's optimal decision rules directly, without having to go through the above long and potentially problematic procedure. This is what duality theory is about. Duality theory is based on the end of the "primal" optimization problem, the optimized objective functions, indirect utility and expenditure functions.

4.2 The indirect utility function and the duality with the utility function

In the primal approach to the study of consumer behavior, the agent chooses a bundle (i.e. a combination of quantities) of goods to maximize the satisfaction generated from its consumption conditional on her budget constraint. As repeated below, a well-behaved utility function is one that satisfies the axioms on preferences laid out before. For the present purposes all axioms are assumed to hold.

From the solution of (1) we obtain the system of uncompensated (Marshallian) demand functions $q(p,I)$. If we substitute them into the objective function, i.e. the utility function, we obtain what is referred to as **Indirect utility function** (IUF). This relationship expresses the highest utility level achievable by the consumer for every set of exogenous variables, prices and income. Thus:

$$V(p,I) = \max_{\{q\}} \left\{ U(q) | p'q \leq I \right\} = U\left[q(p,I) \right] \tag{2}$$

We already noted that the budget constraint will be satisfied as an equality at the optimal point: were it not the case there would exist a bundle generating a higher utility than the "presumed optimal" bundle. This would not be the optimal one. Also, note that unlike $U(q)$, to which here for convenience we will refer as "direct" utility function (DUF), the IUF $V(p,I)$ incorporates an optimiziation process.

There are strict analogies between direct and indirect utility functions, though, beginning with the graphical representation of indifference curves (ICs) in the two good space. We illustrate this point to show how the IUF is derived from the DUF, implying that they are dual to another. If we show that the IUF incorporates all the information concerning the agent's preferences, then one can base the inference directly on the IUF, without having to solve the above cumbersome constrained maximization problem.

Hold money income I constant throughout. We can draw a map of indifference curves in the price space (p_1, p_2) for given I. Every point on a given indifference curve is associated to a pair of prices that generate the same level of (maximum) utility V. These are then level curves for the indirect utility function drawn holding money income constant. These "price" indifference curves are presented on the right-hand side of Figure 4.1, whereas the standard "quantity" indifference curves are portrayed on the left-hand side.

One important difference is that price ICs that are closer to the origin correspond to a higher utility level, which is the opposite relative to the standard case of quantity ICs. Let's compare the set of prices $({p_1}', {p_2}')$ and $({p_1}'', {p_2}'')$. Given the same income I, the consumer must be able to buy more of both goods at lower prices: as a result, she will be more satisfied.

In addition, price ICs are negatively sloped. Indeed, given the pair $({p_1}', {p_2}')$, any other price pair that generates the same level of utility $V(.)$ can not lie above (resp. below) as higher (resp. lower) prices generate a lower (resp. higher) utility. This is shown in the figure's right panel.

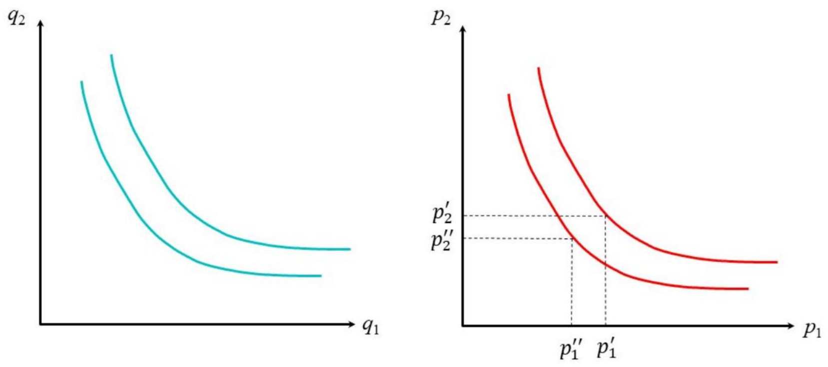

Figure 4.1 *"Quantity" and "Price" indifference curves.*

We now tackle the problem of how to construct the right-hand side IUF-based map of ICs starting from the left-hand side DUF-based map of ICs. Consider Figure 4.2.

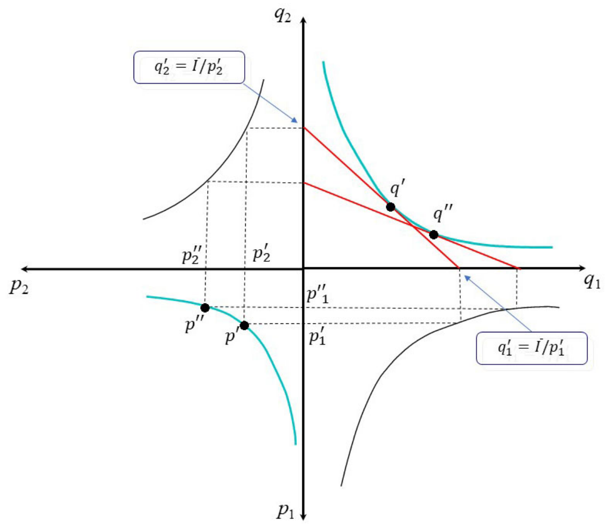

Figure 4.2 *Duality between quantity indifference curves and price indifference curves (1).*

Consider a point such as $q' = (q_1{}', q_2{}')$ and draw a tangent to the IC at that point. It is important to note that this tangent is the budget line for which q' is the optimal choice. We need to find the prices that generate this price line with given income I. These prices are obtained from the intercepts of the price line with the axes. The prices are $p_1{}' = I/q_1{}'$ and $p_2{}' = I/q_2{}'$ because the intercepts are, as we know, given by $q_1{}' = I/p_1{}'$ and $q_2{}' = I/p_2{}'$ respectively. Thus, given income and quantities, we can find the prices. This occurs in the north-east quadrant of the figure. What we need now is to project the prices so found onto the space in the south-west quadrant, i.e. moving from the quantity space to the price space. This is done with the help of the two equilateral hyperbolas whose general representation is $x \cdot y = k$. Letting $k = I$ we have $p_1{}'q_1{}' = I$ and $p_2{}'q_2{}' = I$ which enables us to locate the corresponding price point $p' = (p_1{}', p_2{}')$ in the south-west quadrant. The point has the property that, given income I and prices p', the maximum utility the consumer can achieve is $U(q')$. The process continues by choosing all indifferent points $\widetilde{q} \sim q'$ such as q'' enabling us to construct the price indifference curve corresponding to utility levels equal to $U\left(\widetilde{q}\right)$.

If we repeat the above process for a less preferred point $\widehat{q} \prec q'$ which lies on a lower quantity IC, we come up with a price IC that lies above the previous one and farther away from the origin. This confirms that the closer a price IC to the origin is, the higher is the maximum utility achieved by the consumer selecting the associated bundle, which in turn lies on a far away quantity IC. See Figure 4.3.

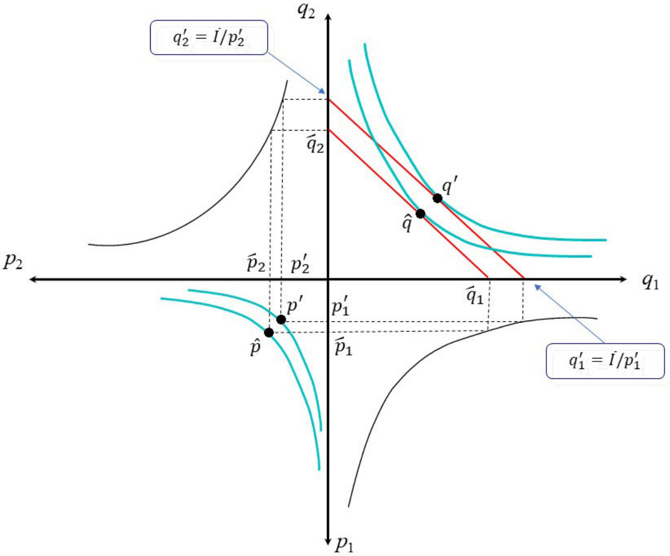

Figure 4.3 *Duality between quantity indifference curves and price indifference curves (2).*

Figure 4.3 can also be read the other way around, enabling going from price ICs to quantity ICs. We start from the point $\widehat{p}$ in the south-west quadrant and obtain the quantities $\widehat{q} = (\widehat{q}_1, \widehat{q}_2)$ which are interpreted as the maximum quantities of the two goods which can be purchased out of income I when the prices are $\widehat{p}_1$ and $\widehat{p}_2$. Hence the line between $\widehat{q}_1$ and $\widehat{q}_2$ is the budget line associated with prices $(\widehat{p}_1, \widehat{p}_2)$. If we now find the IC that is tangent to this budget line, we identify the consumer bundle $(\widehat{q}_1, \widehat{q}_2)$ associated with prices $(\widehat{p}_1, \widehat{p}_2)$. The direct utility $U = (\widehat{q}_1, \widehat{q}_2)$ generated by those quantities, i.e. the utility associated with the IC through $\widehat{q}$ will be the indirect utility function $V = (\widehat{p}_1, \widehat{p}_2, I)$ associated with income I and prices $(\widehat{p}_1, \widehat{p}_2)$. Thus, if $q_1 = (\widehat{p}_1, \widehat{p}_2, I)$ and $q_2 = (\widehat{p}_1, \widehat{p}_2, I)$ are the consumer demand functions, we write $V = \left(\widehat{p}_1, \widehat{p}_2, I\right) = U[q_1\left(\widehat{p}_1, \widehat{p}_2, I\right), q_2\left(\widehat{p}_1, \widehat{p}_2, I\right)]$ which states that the value of the indirect utility function at $\left(\widehat{p}_1, \widehat{p}_2, I\right)$ is equal to the direct utility function evaluated at the demands associated with those prices and income. Hence, to construct $V(.)$ from $U(.)$ we first construct the demand functions and substitute them in $U(.)$ as specified in (2).

4.3 The indirect utility function and its properties

The primal (or Lagrangian) approach to the consumer choice problem utilizes information contained in the direct utility function $U(q)$. As a matter of fact, in many cases we cannot derive specific functions describing the consumer optimal decisions from the direct utility function. The dual approach uses instead information contained in different, but related functions: the indirect utility functions $V(.)$.

If it is not always easy/possible to obtain and analyze the consumer demand functions from the DUF, we may ask whether we can write down a IUF that enables us to derive and analyze the consumer demand functions without having to derive it from the DUF. If the answer is positive, then the analysis is greatly simplified. However, we have to verify if an arbitrary function of prices and income represents a "well-behaved" IUF, i.e. one that satisfies the regularity conditions required for a "valid" indirect utility function.

We therefore proceed to consider these regularity conditions. Note that, generally peaking, three types of properties characterize the central functions of duality theory (besides continuity and differentiability): monotonicity, curvature, and homogeneity.

Properties of the indirect utility function $V(p,I)$:

1. $V(p,I)$ is real-valued, continuous and differentiable with respect to p and I;
2. $V(p,I)$ is non-increasing in p and non-decreasing in I. Thus, for $p' \geq p''$ we have $V(p',I) \leq V(p'',I)$ and for $I' \geq I''$ we have $V(p,I') \geq V(p,I'')$. This implies: $\partial V(p,I)/\partial p_i \leq 0 \;\; \forall i$ and $\partial V(p,I)/\partial I \geq 0$;
3. $V(p,I)$ is homogeneous of degree zero in p and I. Thus $V(\theta p, \theta I) = V(p,I)$ $\theta > 0$;
4. $V(p,I)$ is quasi-convex in p. Every cord in the price space lies above the price indifference curve. That is, for every p' and p'' we have to have $\max\left[V\left(p',I\right), V(p'',I)\right] \geq \left[V(\lambda p' + (1-\lambda)p'', I)\right]$. Thus, the Bordered hessian matrix of $V(p,I)$ relative to p is positive semi-definite. Among other things this implies that $\partial^2 V(p,I)/\partial p_i^2 > 0 \;\; \forall i$.

Proofs.

Monotonicity. a) We know that when the price of one good decreases (resp. increases) the budget line becomes flatter (resp. steeper) so that more (resp. fewer) bundles become affordable. This implies that the new optimal bundle will be located on a higher (resp. lower) indifference curve, and thus higher (resp. lower) maximum utility will be achieved. b) We know that when income increases (resp. decreases) the budget line shifts outward (resp. inward) in a parallel way so that more (resp. fewer) bundles become affordable. This implies that the new optimal bundle will be located on a higher (resp. lower) indifference curve, and thus higher (resp. lower) maximum utility will be achieved. These facts, which we have already seen, are again shown in Figure 4.4.

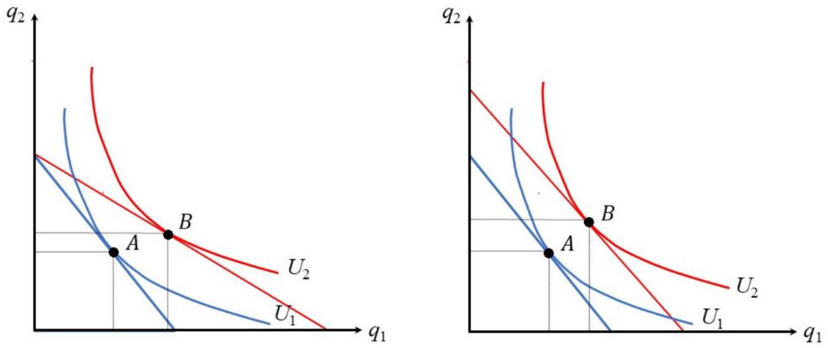

Figure 4.4 *Price and income changes and maximum utility.*

The monotonicity properties of the IUF are represented in Figure 4.5.

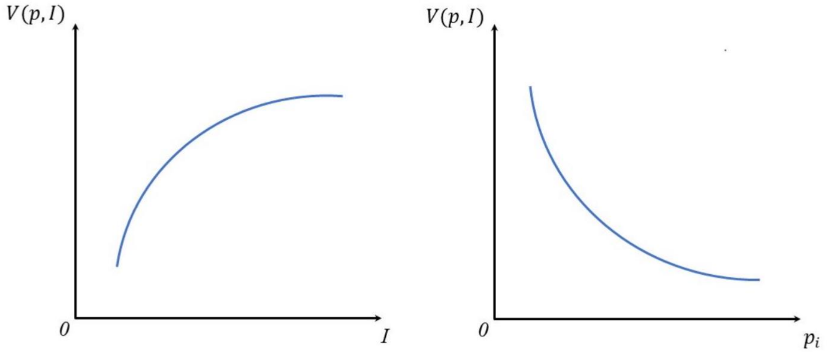

Figure 4.5 *Monotonicity of the indirect utility function.*

Homogeneity. Recall that if all prices and income increase (resp. decrease) by the same proportion θ nothing happens to the budget constraint, that is: $\theta pq = \theta I \implies pq = I$. The position of the budget line in terms of slope and intecepts remains unchanged and therefore the new optimal bundle is the same as the original one. Indeed: $MRS = p_i/p_j = \theta p_i/\theta p_j\ \forall i,j$. As the optimal choice does not change, neither does the maximum utility attained by the agent $U\left(q^*\right) = V(.)$. Note that we can define $\theta = 1/I$, so that $V(p/I,1) = V(p/I)$ where prices are normalized by income. Price indifference curves may have p/I on the axes. See Figure 4.6.

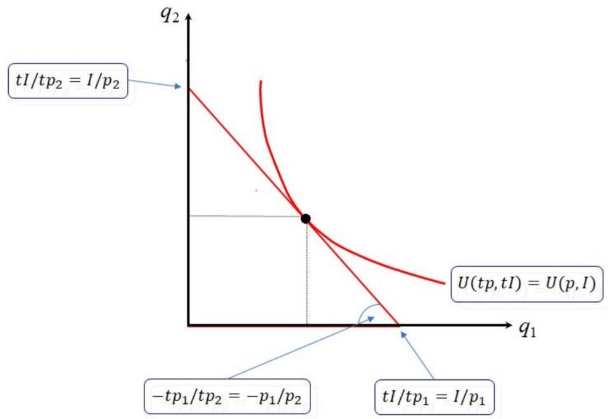

Figure 4.6 *Equal price and income changes and maximum utility.*

Quasi-convexity. This property has been automatically assumed in Figure 4.1. In fact, a function $f(.)$ is quasi-convex if $-f(.)$ is quasi-concave. See Figure 4.5 right panel. For a quasi-convex function its lower contour set is convex. We want to show that if $V(p,I) \geq V(p',I')$, then the indirect utility of the convex combination "budget" is worse than the indirect utility of the (p,I) budget: for any $0 < \lambda < 1$, $V\left[(\lambda p + (1-\lambda)p, \lambda I + (1-\lambda)I)\right] \leq V(p,I)$. The key is that anything in the convex combination "budget" can be afforded with one or the other of the original "budgets". Therefore, at least one of these two "budgets" is at least as good as the convex combination "budget". So, the convex combination is worse than the better one. To show this let $\overline{p} = \lambda p + (1-\lambda)p'$ and $\overline{I} = \lambda I + (1-\lambda)I'$. Also, let $\overline{q} = q(\overline{p}, \overline{I})$. Then $\overline{pq} \leq \overline{I}$ is something you can afford with income $\overline{I}$. This means that $\lambda p\overline{q} + (1-\lambda)p'\overline{q} \leq \lambda I + (1-\lambda)I'$ which in turn implies that either $I - p\overline{q} \geq 0$ or $I' - p'\overline{q} \geq 0$ (or possibly both). The implication is that at least one of the following two things is true: $p\overline{q} \leq I = pq$ and hence $V(\overline{p}, \overline{I}) \leq V(p,I)$ or $p'\overline{q} \leq I' = p', I'$ and hence $V(\overline{p}, \overline{I}) \leq V(p', I')$. So it must be that $V(\overline{p}, \overline{I}) \leq max\left[V(p,I), V(p',I')\right]$.

4.4 Duality between utility function and indirect utility function

If the IUF satisfies the above properties, then duality theory tells us that $V(p,I)$ contains all the information on consumer preferences provided by the DUF $U(q)$. We can informally graphically show this statement using indifference curves. Note that convexity is crucial for the one-to-one relationship between IUF and DUF (otherwise the IUF would exist, but it would not contain "all" information in the DUF). Graphically, we proceed as done above in Figure 4.2 but in the opposite direction. See Figure 4.7.

By varying prices, each point on the price indifference curve generates a quantity indifference curve. The set of budget lines defines the consumption set which lies above each indifference curve. If the IUF is well-behaved, so will be the con-

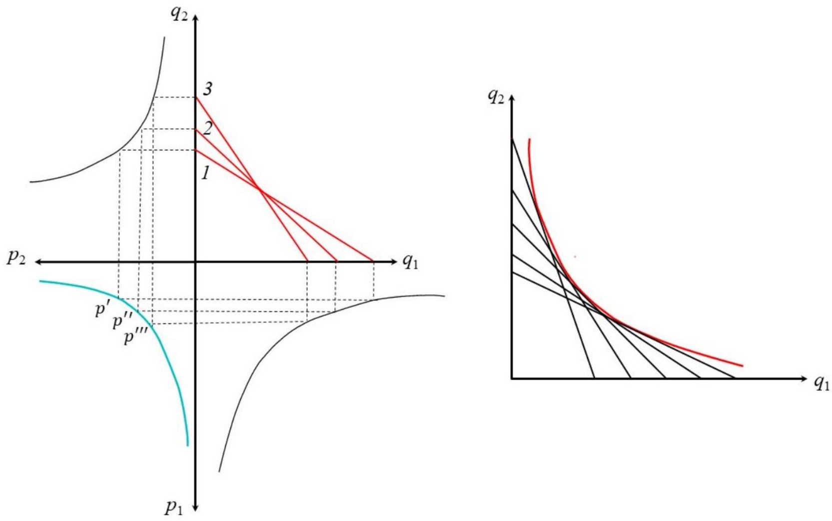

Figure 4.7 *Duality between direct and indirect utility functions.*

sumption set. We can then go from the IUF to the associated DUF. Duality states that the existence of a real-valued direct utility function that is continuous, positive monotonic and quasi-concave implies and is implied by the existence of a real-valued indirect utility function that is continuous, negative monotonic (in p or p/I) and quasi-convex (in p).

Remark 1. The theory can be generalized in two directions: (i) non-differentiability, (ii) zero values for q or p.

Remark 2. (i) Perfect complements or L-shaped preferences: we know that the DUF is not differentiable in q in those points of greater economic interest. It turns out that to L-shaped quantity indifference curves there correspond linear price indifference curves, so that $V(p, I)$ is everywhere differentiable in p. (ii) Perfect substitutes or linear preferences: while the DUF is differentiable in q and quantity indifference curves are linear, the IUF is non differentiable in p at the kink and price indifference curves are L-shaped.

4.5 Duality results using indirect utility functions: Roy's identity and properties of Marshallian demand functions

Formally, the "conjugate" problem to select the bundle that $\max_q \left\{U(q)|p'q \le I\right\}$ is to $\min_p \left\{V(p,I)|p'q \le I\right\}$. Since $V(.)$ is non-increasing in prices, the lowest set of prices which satisfy the budget constraint (note: q given) make utility maximum. Thus:

$$U\left(q^*\right) = \min_p \left\{V(p,I)|p'q \le I\right\} = V\left[p(q), I\right] \tag{3}$$

This is illustrated in Figure 4.8.

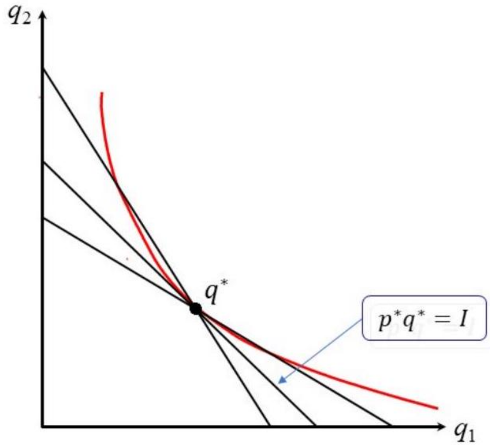

Figure 4.8 *Dual problem: minimization of indirect utility.*

Given the bundle q^* problem (3) invites to consider all the price lines going through q^* and select the one for which q^* is the optimal bundle. The selected line is $p^*q^* = I$ where tangency to the indifference curve obtains. Hence, we are choosing the price that "minimizes the maximum utility attainable". Of course this depends on q.

Let's set up the Lagrangian of problem (3) and compute the FOCs:

$$\Im = V(p, I) - \vartheta(p'q - I) \tag{4}$$

$$V_i(.)/V_j(.) = q_i/q_j \tag{5}$$

Solve (5) for q_i and substitute the result into the budget constraint to get $\sum_{i=1}^n p_i q_i = \sum_{i=1}^n p_i V_i q_j / V_j = I$. Solving for $q_j/V_j = I/\sum_{i=1}^n p_i V_i = 1/\sum_{i=1}^n (p_i/I)V_i$. From this fact and from (5) we get:

$$q_i(p, I) = \frac{q_j(p, I)V_i(p, I)}{V_j(p, I)} = \frac{V_i(p, I)}{\sum_{i=1}^n \left(\frac{p_i}{I}\right) V_i} \tag{6}$$

Recalling that $V(.)$ is homogenous of degree zero in (p, I), Euler's theorem implies that $\sum_{i=1}^n V_i(.)p_i + V_I(.)I = 0$ from which $-V_I(.) = \sum_{i=1}^n V_i(.)(p_i/I)$. Using this fact in (6) we finally get:

$$q_i(p, I) = -\frac{\partial V(p, I)/\partial p_i}{\partial V(p, I)/\partial I} \quad \forall i \tag{7}$$

This is **Roy's identity** and it shows that the full set of Marshallian demand functions can be obtained simply and immediately from differentiating once the indirect utility function with respect to all prices and income.

Proof. An alternative proof of Roy's identity implies invoking the envelope theorem. Recall the constrained utility maximization problem: $\mathscr{L}^* = U(q^*) + \lambda(I - p'q^*)$. From the envelope theorem we know that:

$$\partial\mathscr{L}^*/\partial I = \lambda \quad \text{and} \quad \partial\mathscr{L}^*/\partial p_i = -\lambda q_i^*$$

Because the maximized Lagrangian is equal to the maximized utility (as the constraint at the optimum is equal to zero), then we write:

$$\partial V/\partial I = \lambda \quad \text{and} \quad \partial V/\partial p_i = -\lambda q_i^*$$

Taking the ratio of these two effects yields (7).

We may ask at this point which restrictions on the Marshallian or uncompensated demand functions originate from the duality between direct and indirect utility functions and its main result, Roy's identity. The answer is: not much. Let's see.

Properties of Marshallian demand functions $q(p, I)$.

1. Marshallian demand functions are homogeneous of degree zero in (p, I): for $t > 0$ $q(tp, tI) = q(p, I) = q\left(\widetilde{p}\right)$ where $\widetilde{p} = p/I$ letting $t = 1/I$.

   Proof. Multiplying by $t > 0$ the budget constraint, it remains unchanged and so does the optimal bundle. Hence the utility maximing commoditiy levels do not change when all prices and money income change proportionally by the same amount. Indeed, given:

   $$\begin{cases} \max U(q_1, q_2) \\ s.t. \quad p_1 q_1 + p_2 q_2 = I \end{cases} \Rightarrow \quad \text{FOCs} \Rightarrow \frac{U_1}{U_2} = \frac{p_1}{p_2} \Rightarrow \begin{cases} q_1 = q_1(p_1, p_2, I) \\ q_2 = q_2(p_1, p_2, I) \end{cases}$$

   We now multiply prices and income by $t > 0$ and the problem becomes:

   $$\begin{cases} \max U(q_1, q_2) \\ s.t. \quad tp_1 q_1 + tp_2 q_2 = tI \end{cases} \Rightarrow \quad \text{FOCs} \Rightarrow \frac{U_1}{U_2} = \frac{tp_1}{tp_2} \Rightarrow \begin{cases} q_1 = q_1(p_1, p_2, I) \\ q_2 = q_2(p_1, p_2, I) \end{cases}$$

   These two sets of demand functions are identical.

2. The demand curves are invariant to monotonic transformations of the utility function, i.e. $H\left(q_1, q_2\right) = F\left[U(q_1, q_2)\right]$ where $F' > 0$. That is, the demand functions implied by:

   $$\begin{cases} \max U(q_1, q_2) \\ s.t. \quad p_1 q_1 + p_2 q_2 = I \end{cases}$$

   are identical to those implied by:

   $$\begin{cases} \max H(q_1, q_2) = F\left[U(q_1, q_2)\right] \quad where \quad F' > 0 \\ s.t. \quad p_1 q_1 + p_2 q_2 = I \end{cases}$$

Proof. The first order conditions for the two problems are identical. So are the Marshallian demand functions. That is:

$$\text{FOCs} \Rightarrow \frac{H_1}{H_2} = \frac{F'U_1}{F'U_2} = \frac{U_1}{U_2} = \frac{p_1}{p_2} \Rightarrow \begin{cases} q_1 = q_1(p_1, p_2, I) \\ q_2 = q_2(p_1, p_2, I) \end{cases}$$

There are additional properties of Marshallian demand functions that rely on various types of aggregation. These are important internal checks of the coherency of the theory. The following properties derive from the budget constraint: if empirically they are not satisfied, then the budget constraint is not satisfied. Utility theory is not rejected.

3. Adding up: Marshallian demand functions must satisfy the budget constraint: $\sum_{i=1}^{n} p_i q_i(p, I) = I$.

4. Engle aggregation: $\sum_{i=1}^{n} p_i(\partial q_i / \partial I) = 1$. As the budget constraint has to be always satisfied, the sum of the value of the marginal propensities to consume out of income is equal to one. An elasticity-based alternative is $\sum_{i=1}^{n} \omega_i \eta_i = 1$, i.e. the sum of weighted income elasticities, with weights equal to the share of income spent in good i, equals one.

   Proof. Given the budget constraint $\sum_{i=1}^{n} p_i q_i(p, I) = I$ we differentiate with respect to income to get: $\sum_{i=1}^{n} p_i(\partial q_i / \partial I) = dI/dI = 1$. By multiplying and dividing by (I/q_i) we get the elasticity version.

5. Cournot aggregation: $\sum_{i=1}^{n} p_i(\partial q_i / \partial p_j) + q_j = 0$. Equivalently: $\sum_{i=1}^{n} \omega_i \epsilon_{ij} = -\omega_j$. The sum of all (own and cross) price elasticities of the demand for a good is equal to the negative of its income elasticity.

   Proof. Consider the budget constraint: $\sum_{i=1}^{n} p_i q_i(p, I) = I$. Differentiate each demand function with respect to p_j and the result follows. As to the elasticity version, multiply all terms by (p_j/I) to get: $\sum_{i=1}^{n} p_i(\partial q_i/\partial p_j)(p_j/I) = -p_j q_j / I$. Multiply and divide the left-hand side by q_i and the result obtains.

4.6 The expenditure function and its properties

Figure 4.9 shows the minimum level of expenditure necessary to attain utility level U° given prices (p_1, p_2). The minimum level of expenditure in this case is $p_1 q_1^* + p_2 q_2^*$. Call this I^*, where (q_1^*, q_2^*) are the optimal levels of consumption, given prices (p_1, p_2) and this level of expenditure. The shaded region in the figure shows all the feasible combinations of (q_1, q_2) that will generate $U(q_1, q_2) \geq U^\circ$. Any lower level of expenditure than $p_1 q_1^* + p_2 q_2^*$ will not enable the consumer to attain the utility level U°. If the prices had been different, say (p_1', p_2'), then we would require a different level of expenditure to achieve utility level U°. In this case we would require $p_1{}'q_1{}' + p_2{}'q_2{}'$ (call this level of expenditure I'). Putting this in functional form we write $E\left(p_1, p_2, U^\circ\right)$ as the minimum level of expenditure needed to obtain utility level U° given prices (p_1, p_2). The figure shows $I^* = E\left(p_1, p_2, U^\circ\right)$ and $I' = E\left(p_1{}', p_2{}', U^\circ\right)$.

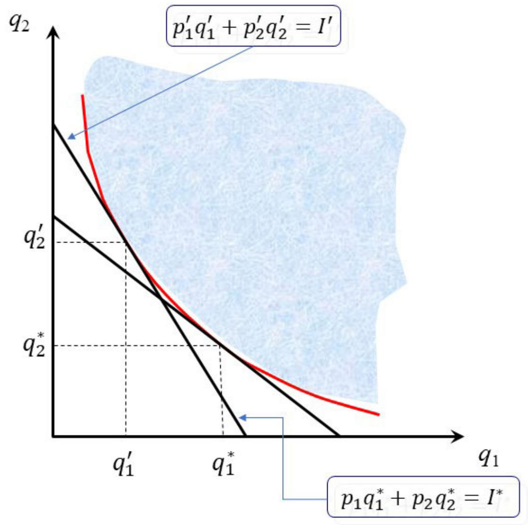

Figure 4.9 *Expenditure minimization.*

Formally, given $p \in R^n_{++}$, q^* solves the problem:

$$E(p, U) = \min_q \left\{p'q | U(q) \geq U\right\} \tag{8}$$

In Figure 4.9 $U = U^\circ$. Note that at the optimal point $E(.) = I$. The figure shows how to derive the expenditure function from the utility function. Given the level U° and $\left(p_1, p_2\right)$ we have to find the bundle (q_1^*, q_2^*) whose minimum expenditure is given by $I^* = p_1 q_1^* + p_2 q_2^*$. Problem (8) yields functions $q = q(p, U)$, called compensated or Hicksian demand functions, from which it follows that the **Expenditure function** is:

$$E(p, U) = p'q = p'q(p, U) \tag{9}$$

The expenditure function is very important in welfare economics (consumer surplus, compensating and equivalent variations) and in index number theory.

Properties of the expenditure function $E(p, U)$:

1. $E(p, U)$ is real-valued, continuous and differentiable with respect to p and U
2. $E(p, U)$ is non-decreasing in p and for all $p > 0$ it is strictly increasing in U. Thus, $\partial E(p, U)/\partial p_i \geq 0$ and $\partial E(p, U)/\partial U > 0$;

3. $E(p,U)$ is homogeneous of degree one in p. Thus, $E(tp,U) = tE(p,U) \quad t > 0$;

4. $E(p,U)$ is concave in p. Thus, the Hessian matrix of $E(p,U)$ is negative semi-definite. Among other things this implies that $\partial^2 E(p,U)/\partial p_i^2 < 0 \quad \forall i$.

Proofs.

Monotonicity. a) Let $p^1 > p^0$. We have: $E(p^1,U) = \min_q \{p^1 q : U(q) \geq U\} = p^1 q \geq p^0 q = \min_q \{p^0 q : U(q) \geq U\} = E(p^0,U)$; b) Let $U^1 > U^0$. We have: $E(p,U^1) = \min_q \{pq : U(q) \geq U^1\} = pq(p,U^1) \geq pq(p,U^0) = \min_q \{pq : U(q) \geq U^0\} = E(p,U^0)$. Basically, this property just means that if you want to attain a higher utility, you have to consume more of the goods, and as long as goods are not free, this means that you will incur a higher expenditure.

Homogeneity. Consider the two expenditure minimization problems: a) $\min_q\{pq : U(q) \geq U\}$ and b) $\min_q \{tpq : U(q) \geq U\}$ where $t > 0$. The FOCs are the same since, for $\forall i,j$: $\frac{U_i}{U_j} = \frac{p_i}{p_j}$ and $\frac{U_i}{U_j} = \frac{tp_i}{tp_j}$. The system of compensated demand functions $q(p,U)$ are the same for the two problems. Thus, we substitute them into the objective function of the first problem to get: $pq(p,U) = E(p,U)$ and into the objective function of the second problem to obtain: $tpq(p,U) = tE(p,U)$. Which proves the result. Diagrammatically, let us consider the familiar illustration of the consumer expenditure minimizing choice of goods in Figure 4.10.

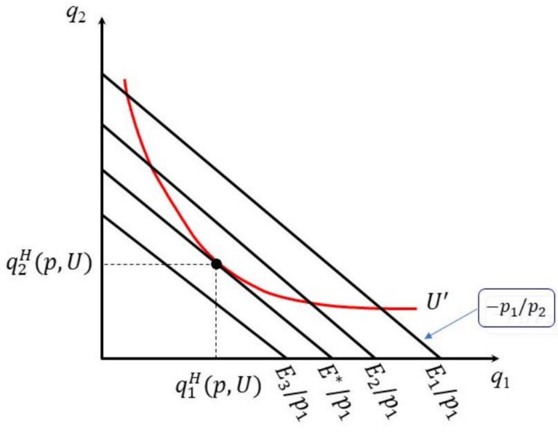

Figure 4.10 *Expenditure minimization.*

It can be seen that if we multiply by the same positive value t all prices, the tangency point denoting the optimal choice does not change. Hence, optimal quantities are unchanged. Also, the position of the tangent isocost line remains the same as its slope is the unchanged goods price ratio and its intercepts with the axes do not change. Indeed, consider the equation of the isocost line when goods prices are multiplied by t: $q_1 = \frac{tE}{tp_1} - \frac{tp_2}{tp_1}q_2 = \frac{E}{p_1} - \frac{p_2}{p_1}q_2$. We see that the isocost line corresponding to the optimal choice must have all goods prices *and* total expenditure multiplied by the same positive value t. This implies that the expenditure function is linear homogeneous.

Concavity. Consider two goods price vectors (p^0, p^1) and their weighted average $\overline{p} = \theta p^0 + (1-\theta)p^1$ with $0 \le \theta \le 1$. Similarly, given the two vectors of goods quantities (q^0, q^1) let $\overline{q} = \theta q^0 + (1-\theta)q^1$. Then, at prices p^0 we have $p^0\overline{q} \ge p^0 q^0 = E\left(p^0, U\right)$. At prices p^1 we have $p^1\overline{q} \ge p^1 q^1 = E\left(p^1, U\right)$. At prices $\overline{p}$ we have $\overline{pq} = \left(\theta p^0 + (1-\theta)p^1\right)\overline{q} = \theta p^0\overline{q} + (1-\theta)p^1\overline{q} = E\left(\overline{p}, U\right)$. Thus: $\overline{pq} = E\left(\overline{p}, U\right) \ge \theta E\left(p^0, U\right) + (1-\theta)E\left(p^1, U\right)$, which proves the (weak) concavity in prices. We can illustrate this property with the help of Figure 4.11.

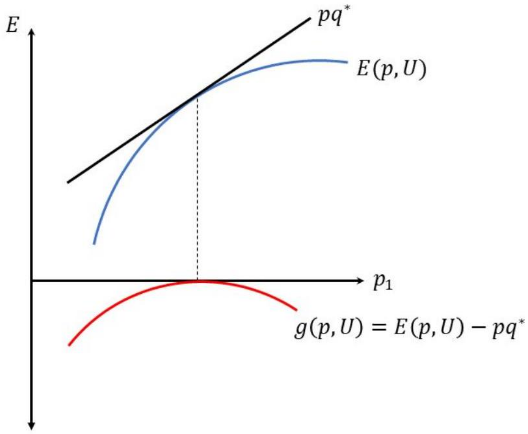

Figure 4.11 *Concavity of expenditure function with respect to goods prices.*

Suppose you graph $E(p, U)$ with respect to p_1 holding (p_2, U) constant. If the good price p_1 rises, the expenditure will never go down as $\partial E(p, U)/\partial p_i \ge 0$ but it increases at a decreasing rate. Why? Because as this one good becomes more expensive and other prices stay the same, there will be some substitution of the more expensive good with the other goods which have become relatively cheaper. Note that in the figure the straight line implies no substitution between goods, whose quantity is not the expenditure minimizing one.

4.7 Duality between utility and expenditure functions

To prove the duality between utility and expenditure functions we have to recall the properties of the former. Thus:

Properties of the utility function $U(q)$, which we will denote synthetically $\boxed{\boldsymbol{U}}$:

1. $U(q)$ is real-valued, continuous and differentiable defined over $q \ge 0$;

2. $U(q)$ is increasing in q. Thus, $\partial U(q)/\partial q_i \ge 0$;

3. $U(q)$ is quasi-concave. Thus, the Bordered hessian matrix of $U(q)$ is negative semi-definite.

Properties of the cost function $E(p,U)$, which we will denote synthetically $\boxed{\boldsymbol{E}}$:

1. $E(p,U)$ is real-valued, continuous and differentiable with respect to p and U;

2. $E(p,U)$is non-decreasing in p and for all $p > 0$ it is strictly increasing in U. Thus: $\partial E(p,U)/\partial p_i \geq 0$ and $\partial E(p,U)/\partial U > 0$;

3. $E(p,U)$ is homogeneous of degree one in p. Thus $E(tp,U) = tE(p,U) \quad t > 0$;

4. $E(p,U)$ is concave in p. Thus, the Hessian matrix of $E(p,U)$ is negative semi-definite. Among other things this implies that $\partial^2 E(p,U)/\partial p_i^2 < 0 \quad \forall i$.

5. $E(p,U)$ is such that the function $U^*(q) = \max\limits_U \left\{U : pq \geq E(p,U)\right\}$ is continuous for $q \geq 0$ and $\forall\ p \geq 0$.

We now have:

Theorem 1. If $U(q)$ satisfies Conditions $\boxed{\boldsymbol{U}}$, then the function

$$E(p,U) = \min_q \left\{pq : U(q) \geq U\right\}$$

satisfies Conditions $\boxed{\boldsymbol{E}}$.

Corollary 1. For every $q \geq 0$ we have $U^*(q) \equiv U(q)$, where $U^*(q)$ is the function defined in Condition n.5 of $\boxed{\boldsymbol{E}}$.

Theorem 2. If $E(p,U)$ satisfies Conditions $\boxed{\boldsymbol{E}}$, then the function $U^*(q)$ satisfies Conditions $\boxed{\boldsymbol{U}}$.

Corollary 2. If the function $E^*(p,U) = \min_q \left\{pq : U^*(q) \geq U\right\}$ then we have $E^*(p,U) \equiv E(p,U)$.

The theorems state that an expenditure function with the right properties corresponds uniquely to a well-behaved utility function. By well-behaved function we mean a function that satisfies the desired properties. The duality theorems together show that an expenditure function can completely describe the preferences of the consumer.

Let us consider a proof of the two duality theorems. Consider the following figure.

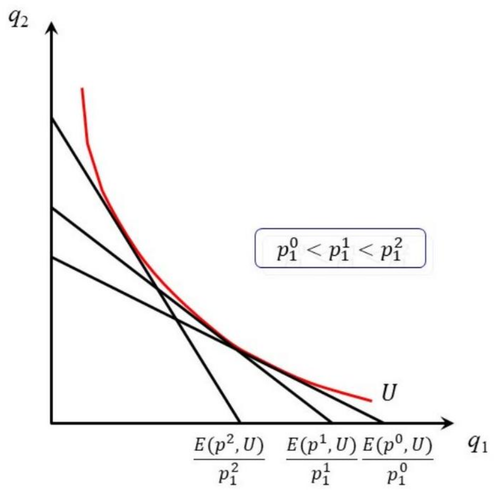

Figure 4.12 *Construction of a well-behaved expenditure function.*

Under Conditions $\boxed{\boldsymbol{U}}$ the consumption set is regular, monotonic, and convex. The associated isoquant has the shape in the figure. For each alternative vector $p = (p_1, p_2)$ – in the graph three vectors are shown – compute $E(p,U) = \min_q \{pq : U(q) \geq U\}$. The expenditure function is built in this way and can be shown to satisfy Conditions $\boxed{\boldsymbol{E}}$.

Now reverse the argument. Suppose we know $E(p,U)$. To find the consumption set $C^*(U^0)$ defined as $C^*(U) = \{q \in R_+ : pq \geq E(p,U), \forall p \geq 0\}$, we choose a vector $p^0 = ({p_1}^0, {p_2}^0)$, compute $E(p^0, U)$ and eliminate all $q : p^0 q < E(p^0, U)$. Repeat the same procedure for alternative p's and the shaded area in the figure remains. That area corresponds to a well-behaved input requirement set $C^*(U)$.

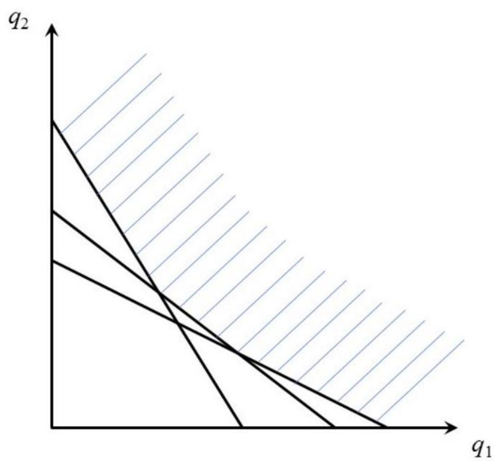

Figure 4.13 *Construction of a well-behaved consumption set.*

The powerful implication of duality is that the expenditure function summarizes all the economically relevant aspects of the consumer's utility and preferences.

4.8 Duality results using expenditure functions: Shephard's lemma and properties of Hicksian demand functions

The main result of the duality between utility and expenditure (or cost) functions is known as Shephard's lemma. This is a major result in Microeconomics having applications not only in consumer theory but also in producer theory. Shephard's lemma enables to obtain the whole set of Hicksian demand functions by simple differentiation of the expenditure function with respect to goods prices. That is:

Shepard's lemma. Letting $q_i(p,U)$ be the compensated or Hicksian demand function for good i, then:

$$q_i(p,U) = \frac{\partial E(p,U)}{\partial p_i} \quad \forall i \tag{10}$$

Proof. Alternative proofs.

Proof 1. Given the initial situation $\left(p^*, q^*\right)$. By definition we have $E\left(p^*, U\right) = \sum_{i=1}^{n} p_i^* q_i^*$. We let the price of good j change (with $dU = dp_k = 0\ \forall k \neq j$): $dE(.) = \sum_i p_i^* dq_i^* + q_j^* dp_j^*$. Now note: (i) $dU = \sum_i U_i dq_i = 0$, (ii) $p_i = \mu U_i(.)$ from the FOCs of expenditure minimization. Then: $\sum_i p_i dq_i = \sum_i \mu U_i(.)\ dq_i = \mu \sum_i U_i(.)\ dq_i = 0$. It follows that $dE(.) = q_j^* dp_j^*$. Hence the result.

Proof 2. Envelope theorem. The problem is to $\min_q p'q + \mu\left[U(q) - U\right]$. From the FOCs we obtain the Hicksian demand functions $q = q(p,U)$. The optimized Lagrangian is $\mathscr{L}^* = pq(p,U) + \mu\left\{U\left[q(p,U)\right] - U\right\}$. We are interested in the impact of a change in the price of a good: $\frac{\partial \mathscr{L}^*}{\partial p_j} = q_j(p,U) + \sum_i \left(p_i \frac{\partial q_i}{\partial p_j} - \mu \frac{\partial U}{\partial q_i}\frac{\partial q_i}{\partial p_j}\right) = q_j(p,U) + \sum_i \left(p_i - \mu \frac{\partial U}{\partial q_i}\right)\frac{\partial q_i}{\partial p_j} = q_j(p,U)$. Note that the term in round brackets is zero because of the FOCs. Racalling that at the optimum the constraint is zero, then $\frac{\partial \mathscr{L}^*}{\partial p_j} = \frac{\partial E(p,U)}{\partial p_j}$. Hence the result.

Proof 3. Distance formula (see Figure 4.11). Let q^* be the cost minimizing bundle and $E(p,U) = pq^*$. Define the function: $g(p,U) \equiv E(p,U) - pq^*$. Clearly $g(p,U) \leq 0$ and $g(p,U) = 0$ at the maximum. The necessary condition for a maximum is $\frac{\partial g(p,U)}{\partial p_i} = \frac{\partial E(p,U)}{\partial p_i} - q_i^* = 0$.

Properties of compensated or Hicksian demand functions $q(p,U)$:

1. Substitution effect: Hicksian demand functions are inversely related to their own price: $\frac{\partial q_i(p,y)}{\partial p_i} \leq 0\ \forall i$. This implies that demand curves are downward sloped.

Proof. By the concavity of the expenditure function and by Shephard's lemma: $\frac{\partial q_i(p,U)}{\partial p_i} = \frac{\partial^2 E(p,U)}{\partial p_i^2} \leq 0$.

2. Hicksian demand functions are homogeneous of degree zero in prices p: $q(tp,U) = q(p,U)$ $t > 0$.

   Proof. By the properties of homogeneous functions, because the expenditure function is O^1 in p, then its first partials are O^0 in p.

3. Cross price effects are symmetric: $\frac{\partial q_i(p,U)}{\partial p_j} = \frac{\partial q_j(p,U)}{\partial p_i}$.

   Proof. By the symmetry of the expenditure function and by Shephard's lemma: $\frac{\partial q_i(p,U)}{\partial p_j} = \frac{\partial^2 E(p,U)}{\partial p_i \partial p_j} = \frac{\partial^2 E(p,U)}{\partial p_j \partial p_i} = \frac{\partial q_j(p,U)}{\partial p_i}$.

There are additional properties of compensated demand functions that rely on various types of aggregation. These are important internal checks of the coherency of the theory.

4. The following holds: $\sum_{j=1}^{n} \frac{\partial q_j(p,U)}{\partial p_j} p_j = 0$. The sum of price responses of demand for good i, weighted by prices, is equal to zero.

   Proof. The result follows from the fact that conditional input demand functions are O^0 in p. Apply Euler's theorem and you get the results.

4.9 Relationship between indirect utility function and expenditure functions

Having seen how the EF and the IUF are derived from the direct utility function there must obviously be some relationship between them. To see this consider Figure 4.14 which shows the indifference curve corresponding to the level of utility U° and a budget line $pq = I^\circ$. Levels U° and I° are picked in such a way that the budget line is tangential to the indifference curve at q^*. Clearly q^* maximizes utility subject to the constraint $pq \leq I^\circ$. Put another way this tells us that U° is the maximum utility given the budget constraint $pq \leq I^\circ$, i.e. $V\left(p, I^\circ\right) = U^\circ$.

It is also obvious from the figure that the minimum expenditure necessary to achieve utility U°, given prices p, is I°, i.e. $E\left(p, U^\circ\right) = I^\circ$. Since both these relationships between U° and I° must be true, we have:

$$E\left(p, U^\circ\right) = I^\circ \Longleftrightarrow V\left(p, I^\circ\right) = U^\circ \tag{11}$$

Because the expenditure function is increasing in utility, we can invert it and solve it for U to obtain the indirect utility function: $U = E^{-1}(p, I) = V(p, I)$. Similarly, since the indirect utility function is increasing in income, we can invert it and solve it for I to obtain the expenditure function: $I = V^{-1}(p, U) = E(p, U)$.

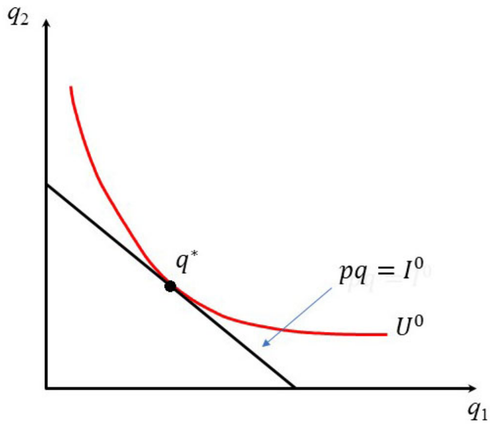

Figure 4.14 *Indirect utility and expenditure functions.*

4.10 Duality theory of consumer behavior: summary

To summarize, we visualize the duality existing between direct utility function and expenditure function on the one hand and between direct utility and indirect utility functions on the other hand. These two functions are in turn related as shown in (11). That is:

$$\begin{aligned} E(p,U) &\equiv E\left[p, V(p,I)\right] \equiv I \\ V(p,I) &\equiv V\left[p, E(p,U)\right] = U \end{aligned} \tag{12}$$

The duality for consumer behavior is portrayed in Figure 4.15.

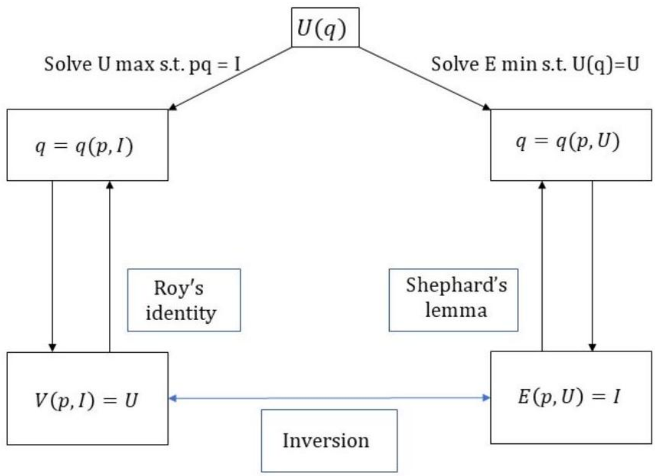

Figure 4.15 *Duality in consumer theory.*

4.11 Duality theory and the Slutsky equation

We now investigate the relationship between Marshallian $q_i^{(M)} = q_i(p, I)$ and Hicksian $q_i^{(H)} = q_i(p, U)$ demand functions. We will exploit duality theory and specifically Shephard's lemma and Roy's identity. Consider the second equation in (12) $V\left[p, E(p, U)\right] = U$ and differentiate it with respect to p_i:

$$\frac{\partial V(p,I)}{\partial p_i} + \frac{\partial V(p,I)}{\partial I}\frac{\partial E(p,U)}{\partial p_i} = 0$$

from which we obtain:

$$q_i(p,I) = -\frac{\partial V(.)}{\partial p_i} / \frac{\partial V(.)}{\partial I} = \frac{\partial E(.)}{\partial p_i} = q_i(p,U) \tag{13}$$

This corresponds graphically to the point where Hicksian and Marshallian demand curves intersect one another: see Figure 4.16.

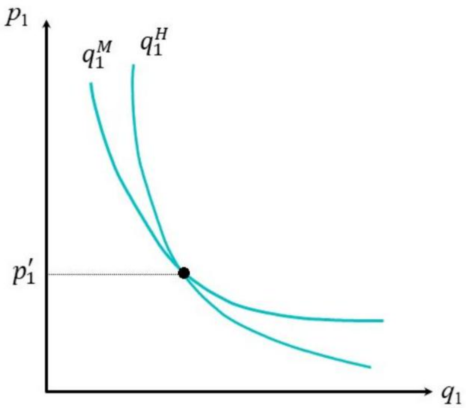

Figure 4.16 *Marshallian and Hicksian demand functions.*

We can rewrite (13) as: $q_i^{(M)} = q_i\left[p, E(p, U)\right] = q_i(p, U) = q_i^{(H)}$. This is an identity valid for all (p, U). We can differentiate the expression with respect to p_j:

$$\frac{\partial q_i(p,I)}{\partial p_j} + \frac{\partial q_i(p,I)}{\partial I}\frac{\partial E(p,U)}{\partial p_j} = \frac{\partial q_i(p,U)}{\partial p_j}$$

Using Shephard's lemma:

$$\frac{\partial q_i(p,I)}{\partial p_j} = \frac{\partial q_i(p,U)}{\partial p_j} - q_j\frac{\partial q_i(p,I)}{\partial I} \quad \forall i, j \tag{14}$$

This is the **Slutsky substitution matrix**. Combined with Cournot and Engel aggregation properties, (14) generates important restrictions on demand functions for goods.

Note also the following result. We ask under which conditions we have symmetry of the Marshallian cross price effects $\frac{\partial q_i(p,I)}{\partial p_j} = \frac{\partial q_j(p,I)}{\partial p_i}$. Using the Slutsky equation and recalling that one of the properties of Hicksian demand functions is the symmetry of cross price effects $\frac{\partial q_i(p,U)}{\partial p_j} = \frac{\partial q_j(p,U)}{\partial p_i}$, we find that Marshallian demand functions exhibit symmetry of cross price effects if and only if $\frac{\partial q_i(p,I)}{\partial I} q_j = \frac{\partial q_j(p,I)}{\partial I} q_i$ or $\frac{\partial q_i(p,I)}{\partial I} \frac{I}{q_i} = \frac{\partial q_j(p,I)}{\partial I} \frac{I}{q_j} \Longrightarrow \eta_i = \eta_j$. The income elasticities have to be the same.

Chapter 5

Producer behavior: duality between technology and costs and between technology and profit

5.1 Duality in producer behavior

Up to this point the firm has been characterized by its production function and from it we have derived cost and profit functions, together with the decision rules in terms of input demand and output supply functions. The question we tackle in this chapter is whether or not we can characterize the firm by means of its cost or profit function on the condition that they reflect the properties of the original production function.

What is not obvious in general is that (under certain regularity conditions) the cost function $C(w,\ y)$ also completely describes the technology of the given firm: i.e., the firm's cost function $C(w,\ y)$ can be used to define the firm's production function $f(x)$. Thus, there is a duality between cost and production functions in the sense that either of them can describe the technology of the firm equally well in certain circumstances. The same holds for the duality between production and profit functions $\Pi(p, w)$.

The strategy we will follow is to derive the regularity conditions that a cost function must have (irrespective of the functional form or specific regularity properties for the production function) and to show how a production function can be constructed from a given cost function. We will then rely on the cost function so validated to obtain the optimal decision rules, i.e. the conditional input demand functions, in a straightforward way. We will do the same for the profit function.

Note that the results below carry over to the case of multiproduct technologies, with p and y being vectors rather than scalars.

5.2 The cost function and its properties

In the standard approach, what we call sometimes the primal approach, to the study of producer behavior, the producer chooses input levels to minimize the total cost subject to the technological constraint represented by a well-behaved production function (or transformation function in a multi-output setting). A well-behaved production function is one that satisfies the properties of the technology set laid out before.

The steps followed in the solution of the cost minimization problem are to compute the FOCs and check the SOCs, to solve the FOCs explicitly for the conditional input demand functions, and to substitute these functions into the objective function to obtain the minimum cost function. That is:

$$C(w,y) = \min_x \left\{wx : y \leq f(x)\right\}$$
$$C(w,y) = wx = w\phi(w,y) \tag{1}$$

where x and w are n-element vectors, while y is taken to be a scalar, although it could be replaced by a m-element vector without any problem. It should be clear that the cost function will be characterized by a number of properties that originate from the properties of the production function and the optimization problem that generated it. We therefore proceed to consider these properties. Note that three types of properties usually characterize the functions of duality theory (besides continuity and differentiability): monotonicity, curvature, and homogeneity.

Properties of the cost function $C(w,y)$:

1. $C(w,y)$ is real-valued, continuous and differentiable with respect to w and y and it is zero when $y=0$;

2. $C(w,y)$ is non-decreasing in w and for all $w>0$ it is strictly increasing in y. Thus: $\partial C(w,y)/\partial w_i \geq 0$ and $\partial C(w,y)/\partial y > 0$;

3. $C(w,y)$ is homogeneous of degree one in w. Thus $C(tw,y) = tC(w,y) \;\; t>0$;

4. $C(w,y)$ is concave in w. Thus, the Hessian matrix of $C(w,y)$ is negative semi-definite. Among other things this implies that $\partial^2 C(w,y)/\partial w_i^2 < 0 \;\; \forall i$.

Proofs.

Monotonicity. a) Let $w^1 > w^0$. We have: $C\left(w^1,y\right) = \min_x \left\{w^1 x : y \leq f(x)\right\} = w^1 x \geq w^\circ x = \min_x \left\{w^\circ x : y \leq f(x)\right\} = C(w^\circ, y)$; b) Let $y^1 > y^\circ$. We have: $C\left(w,y^1\right) = \min_x \left\{wx : y^1 \leq f(x)\right\} = wx\left(w,y^1\right) \geq wx(w,y^\circ) = \min_x \left\{wx : y^\circ \leq f(x)\right\} = C(w,y^\circ)$.

Basically, this property just means that if you want to produce more output, you have to use more inputs, and as long as inputs are not free, this means that you will incur higher cost.

Homogeneity. Consider the two cost minimization problems: a) $\min_x \{wx : y \leq f(x)\}$ and b) $\min_x \{twx : y \leq f(x)\}$ where $t > 0$. The FOCs are the same since: $\frac{f_i}{f_j} = \frac{w_i}{w_j}$ and $\frac{f_i}{f_j} = \frac{tw_i}{tw_j}$ for $\forall i, j$. The system of conditional input demand functions $x(w, y)$ are the same for the two problems. Thus, we substitute them into the objective function of the first problem to get: $wx(w, y) = C(w, y)$ and into the objective function of the second problem to obtain: $twx(w, y) = tC(w, y)$. Which proves the result. Diagrammatically, let us consider the familiar illustration of the producer's cost minimizing choice of inputs in Figure 5.1.

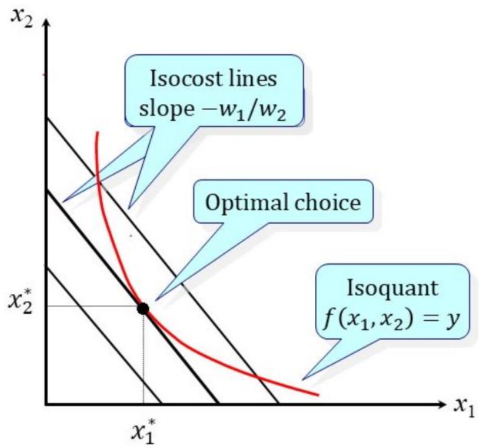

Figure 5.1 *Cost minimization.*

It can be seen that if we multiply by the same positive value t all prices, the tangency point denoting the optimal choice does not change. Hence, optimal input quantities are unchanged. Also, the position of the tangent isocost line remains the same as its slope is the unchanged input price ratio and its intercepts with the axes do not change. Indeed, consider the equation of the isocost line when input prices are multiplied by t: $x_1 = \frac{tC}{tw_1} - \frac{tw_2}{tw_1}x_2 = \frac{C}{w_1} - \frac{w_2}{w_1}x_2$. We see that the isocost line corresponding to the optimal choice must have all inpt prices *and* total cost multiplied by the same positive value t. This implies that the cost function is linear homogeneous.

Concavity. Consider two input price vectors (w^0, w^1) and their weighted average $\overline{w} = \theta w^0 + (1-\theta)w^1$ with $0 \leq \theta \leq 1$. Similarly, given the two vectors of input quantities (x^0, x^1) let $\overline{x} = \theta x^0 + (1-\theta)x^1$. Then, at prices w^0 we have $w^0\overline{x} \geq w^0 x^0 = C\left(w^0, y\right)$. At prices w^1 we have $w^1\overline{x} \geq w^1 x^1 = C\left(w^1, y\right)$. At prices $\overline{w}$ we have $\overline{wx} = \left(\theta w^0 + (1-\theta)w^1\right)\overline{x} = \theta w^0\overline{x} + (1-\theta)w^1\overline{x} = C\left(\overline{w}, y\right)$. Thus: $\overline{wx} = C\left(\overline{w}, y\right) \geq$

$\theta C\left(w^0, y\right) + (1-\theta)C\left(w^1, y\right)$, which proves the (weak) concavity in prices. We can illustrate this property with the help of Figure 5.2.

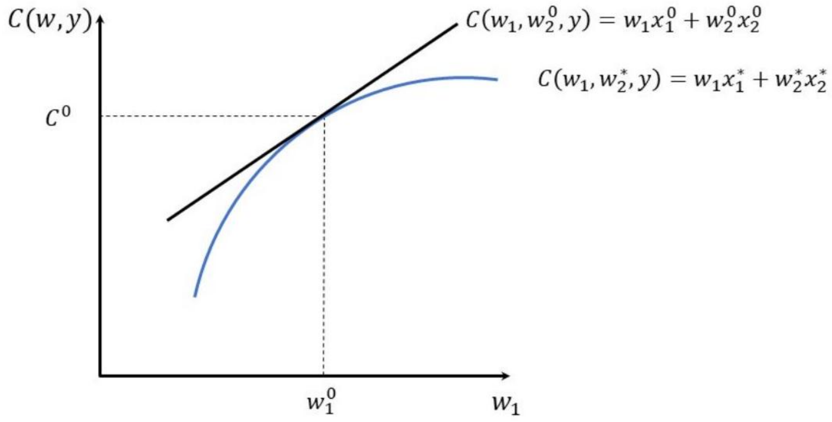

Figure 5.2 *Concavity of the cost function with respect to input prices.*

This property may seem surprising. Suppose you graph $C(w, y)$ with respect to w_1 holding (w_2, y) constant. If the factor price w_1 rises, costs will never go down as $\partial C(w, y)/\partial w_i \geq 0$ but they increase at a decreasing rate. Why? Because as this one factor becomes more expensive and other prices stay the same, there will be some substitution of the more expensive input with the other inputs which have become relatively cheaper. Note that in the figure the straight line implies no substitution between factors, whose quantity is not the cost minimizing one.

5.3 Duality between production and cost functions

To prove the duality between production and cost functions we have to recall the properties of the technology. We can express these properties in terms of production possibility set Y or of input requirement set $V(y)$, or again – in the case of single output technologies – of production functions. We will use this last representation. Thus:

Properties of the production function $f(x)$, which we will denote synthetically $\boxed{\boldsymbol{F}}$:

1. $f(x)$ is real-valued, continuous and differentiable defined over $x \geq 0$;

2. $f(x)$ is nondecreasing in x. Thus: $\partial f(x)/\partial x_i \geq 0$;

3. $f(x)$ is quasi-concave. Thus, the Bordered hessian matrix of $f(x)$ is negative semi-definite.

These properties imply that $V(y)$ is regular, monotonic and convex.

Properties of the cost function $C(w, y)$, which we will denote synthetically $\boxed{\boldsymbol{C}}$:

1. $C(w, y)$ is real-valued, continuous and differentiable with respect to w and y and it is zero when $y = 0$;
2. $C(w, y)$ is non-decreasing in w and for all $w > 0$ it is strictly increasing in y. Thus: $\partial C(w, y)/\partial w_i \geq 0$ and $\partial C(w, y)/\partial y > 0$;
3. $C(w, y)$ is homogeneous of degree one in w. Thus $C(tw, y) = tC(w, y) \;\; t > 0$;
4. $C(w, y)$ is concave in w. Thus, the Hessian matrix of $C(w, y)$ is negative semi-definite. Among other things this implies that $\partial^2 C(w, y)/\partial w_i^2 < 0 \;\; \forall i$.
5. $C(w, y)$ is such that the function $f^*(x) = \max_y \left\{y : wx \geq C(w, y)\right\}$ is continuous for $x \geq 0$ and $\forall \; w \geq 0$.

We now have:

Theorem 1. If $f(x)$ satisfies Conditions $\boxed{\boldsymbol{F}}$, then the function $C(w, y) = \min_x \left\{wx : f(x) \geq y\right\}$ satisfies Conditions $\boxed{\boldsymbol{C}}$.

Corollary 1. For every $x \geq 0$ we have $f^*(x) \equiv f(x)$, where $f^*(x)$ is the function defined in property n.5 of $\boxed{\boldsymbol{C}}$.

Theorem 2. If $C(w, y)$ satisfies Conditions $\boxed{\boldsymbol{C}}$, then the function $f^*(x)$ satisfies Conditions $\boxed{\boldsymbol{F}}$.

Corollary 2. Given the function $C^*(w, y) = \min_x \left\{wx : f^*(x) \geq y\right\}$ we have $C^*(w, y) \equiv C(w, y)$.

The theorems state that a cost function with the right properties corresponds uniquely to a well-behaved production function. By well-behaved function we mean a function that satisfies the desired properties. The duality theorems together show that a cost function can completely describe a well-behaved technology. This is a powerful result that implies that we can base all our inference on the cost function: in McFadden's words "*the cost function is a sufficient statistics for the production function*" (Nobel laureate Daniel Mc Fadden has been a prominent contributor to duality theory).

Let us consider a proof of the two duality theorems. Consider the Figure 5.3.

Under Conditions $\boxed{\boldsymbol{F}}$ $V(y)$ is regular, monotonic, and convex. The associated isoquant $Q(y)$ has the shape in the figure. For each alternative vector $w = (w_1, w_2)$ – in the graph three vectors are shown – compute $C(w, y) = \min_x \left\{wx : f(x) \geq y\right\}$. The cost function is built in this way and can be shown to satisfy Conditions $\boxed{\boldsymbol{C}}$.

Now reverse the argument. Suppose we know $C(w, y)$. To find $V^*(y)$ defined as $V^*(y) = \left\{x \in R_+ : wx \geq C(w, y), \forall w \geq 0\right\}$, we choose a vector $w^\circ = \left(w_1^\circ, w_2^\circ\right)$, compute $C\left(w^\circ, y\right)$ and eliminate all $x : w^\circ x < C\left(w^\circ, y\right)$. Repeat the same procedure

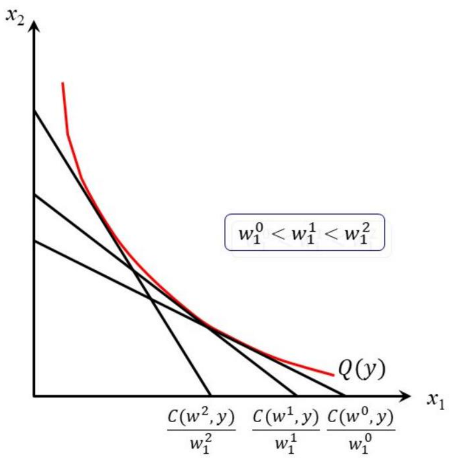

Figure 5.3 *Construction of a well-behaved cost function.*

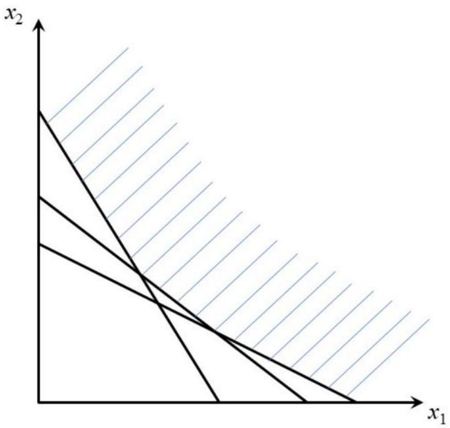

Figure 5.4 *Construction of a well-behaved input requirement set.*

for alternative w's and the shaded area in the Figure 5.4 remains. That area corresponds to a well-behaved input requirement set $V^*(y)$.
The powerful implication of duality is that the cost function summarizes all the economically relevant aspects of the firm's technology.

5.4 Duality results using cost functions: Shephard's lemma and properties of conditional input demand functions

The main result of the duality between production and cost functions is known as Shephard's lemma. This is a major result in Microeconomics having applications not only in the theory of the firm but also in consumer theory. The lemma is named after Ronald Shephard who gave a proof in his 1953 book *Theory of Cost and Production Functions.* The equivalent result in the context of consumer theory was first derived by Lionel W. McKenzie in 1957. Similar results had already been derived by John Hicks in 1939 and Paul Samuelson in 1947.

Shephard's lemma enables to obtain the whole set of conditional input demand functions by simple differentiation of the cost function with respect to input prices. That is:

Shepard's lemma. Let $x_i(w, y)$ be the conditional demand for input i, then:

$$x_i(w, y) = \frac{\partial C(w, y)}{\partial w_i} \quad \forall i \tag{2}$$

Proofs. Although Shephard's original proof used the distance formula, modern proofs of the Shephard's lemma use the envelope theorem. We show both proofs here.

Proof 1. Envelope theorem. The problem is to $\min_x wx + \lambda\left[y - f(x)\right]$. From the FOCs we obtain the conditional input demand functions $x = x(w, y)$. The optimized Lagrangian is $\mathscr{L}^* = wx(w, y) + \lambda\left\{y - f\left[x(w, y)\right]\right\}$. We are interested in the impact of a change in the price of an input: $\frac{\partial \mathscr{L}^*}{\partial w_j} = x_j(w, y) + \sum_i (w_i \frac{\partial x_i}{\partial w_j} - \lambda \frac{\partial f}{\partial x_i}\frac{\partial x_i}{\partial w_j}) = x_j(w, y) + \sum_i (w_i - \lambda \frac{\partial f}{\partial x_i})\frac{\partial x_i}{\partial w_j} = x_j(w, y)$. Note that the term in round brackets is zero because of the FOCs. Recalling that at the optimum the constraint is zero, then $\frac{\partial \mathscr{L}^*}{\partial w_j} = \frac{\partial C(w,y)}{\partial w_j}$. Hence the result.

Proof 2. Distance formula. Let x^* be the cost minimizing bundle and $C(w, y) = wx^*$. Define the function: $g(w, y) \equiv C(w, y) - wx^*$. Clearly $g(w, y) \le 0$ and $g(w, y) = 0$ at the maximum. The necessary condition for a maximum is $\frac{\partial g(w,y)}{\partial w_i} = \frac{\partial C(w,y)}{\partial w_i} - x_i^* = 0$.

Shephard's lemma holds also in consumer theory for well-behaved expenditure functions: differentiation with respect to commodity prices yields the system of Hicksian demand functions. Combined with Shephard's lemma the properties of the cost function entails the following.

Properties of conditional input demand functions $x(w, y)$:

1. Conditional input demand functions are inversely related to their own price: $\frac{\partial x_i(w,y)}{\partial w_i} \le 0 \; \forall i$. This implies that input demand curves are downward sloped.
   Proof. By the concavity of the cost function and by Shephard's lemma: $\frac{\partial x_i(w,y)}{\partial w_i} = \frac{\partial^2 C(w,y)}{\partial w_i^2} \le 0$.
2. Cross price effects are symmetric: $\frac{\partial x_i(w,y)}{\partial w_j} = \frac{\partial x_j(w,y)}{\partial w_i}$.
   Proof. By the symmetry of the cost function and by Shephard's lemma: $\frac{\partial x_i(w,y)}{\partial w_j} = \frac{\partial^2 C(w,y)}{\partial w_i \partial w_j} = \frac{\partial^2 C(w,y)}{\partial w_j \partial w_i} = \frac{\partial x_j(w,y)}{\partial w_i}$.

3. Conditional input demand functions are homogeneous of degree zero in w: $x(tw, y) = x(w, y) \quad t > 0$.
   Proof. By the properties of homogeneous functions, because the cost function is O^1 in w, then its first partials are O^0 in w.

There are additional properties of conditional input demand functions that rely on various types of aggregation. These are important internal checks of the coherency of the theory.

4. The following holds: $\sum_{j=1}^{n} \frac{\partial x_j(w,y)}{\partial w_j} w_j = 0$.
   Proof. The result follows from the fact that conditional input demand functions are O^0 in w. Apply Euler's theorem and you get the results.

5. The following holds: $\sum_{j=1}^{n} \frac{\partial x_j(w,y)}{\partial y} w_j \geq 0$. This implies that not all inputs can be inferior.
   Proof. Since the cost function is O^1 in w apply Euler's theorem to get: $\sum_{i=1}^{n} \frac{\partial C(w,y)}{\partial w_i} w_i = C(w, y)$. Differentiate both terms with respect to y and use Shephard's lemma: $\sum_{i=1}^{n} \frac{\partial C^2(w,y)}{\partial w_i \partial y} w_i = \sum_{i=1}^{n} \frac{\partial x_i(w,y)}{\partial y} w_i = \frac{\partial C(w,y)}{\partial y} \geq 0$. The positive sign follows from the monotonicity property of the cost function with respect to output.

6. The sum of all price elasticities of the conditional demand function for an individual input is zero: $\sum_{j=1}^{n} \epsilon_{ij} = 0$.
   Proof. Since the conditional input demand function is O^0 in w apply Euler's theorem to get: $\sum_{j=1}^{n} \frac{\partial x_i(w,y)}{\partial w_j} w_j = 0$. Divide each term by x_i to get: $\sum_{j=1}^{n} \frac{\partial x_i(w,y)}{\partial w_j} \frac{w_j}{x_i} = 0$. Hence the result.

7. An appropriately weighted sum of output elasticities of all conditional input demand functions is equal to one: $\sum_{i=1}^{n} \alpha_i \epsilon_{iy} = 1$ where $\epsilon_{iy} = \frac{\partial x_i(w,y)}{\partial y} \frac{y}{x_i}$ and $\alpha_i = \frac{w_i x_i}{\lambda y}$ (note that λ is marginal cost and under price-taking profit maximization is equal to the output price. In this case α_i is the ratio between the cost of input i and total revenue).
   Proof. Take the production function and totally differentiate it. Then divide throughout by y to get: $1 = \frac{dy}{dy} = \sum_{i=1}^{n} f_i(x) \frac{\partial x_i}{\partial y}$. Use the FOCs of cost minimization $w_i = \lambda f_i(x)$ to obtain: $1 = \sum_{i=1}^{n} \frac{w_i}{\lambda} \frac{\partial x_i}{\partial y}$. Multiply and divide the right-hand side by x_i / y to get: $1 = \sum_{i=1}^{n} \left(\frac{w_i x_i}{\lambda y} \right) \frac{\partial x_i}{\partial y} \frac{y}{x_i}$. Hence the result.

5.5 The profit function and its properties

In the primal approach to the study of producer behavior, the producer chooses input levels to maximize profits subject to the technological constraint represented by a well-behaved production function (or transformation function in a multi-output setting). Recall that a well-behaved production function is one that satisfies the properties of the technology set laid out before.

The steps followed in the solution of the profit maximization problem are to compute the FOCs and check the SOCs, to solve the FOCs explicitly for the unconditional input demand functions, and to substitute these functions into the technological constraint to obtain the output supply functions and into the objective function to obtain the maximum profit function. That is:

$$\Pi(w,p) = \max_x \{pf(x) - wx\}$$

$$\Pi(w,p) = pf\left[\psi(w,p)\right] - w\psi(w,p) \tag{3}$$

where x and w are n-element vectors, while $y = f(x)$ is taken to be a scalar, although it could be replaced by a m-element vector without any problem. It should be clear that the profit function will be characterized by a number of properties that originate from the properties of the production function and the optimization problem that generated it. We therefore proceed to consider these properties. Note that three types of properties usually characterize the functions of duality theory (besides continuity and differentiability): monotonicity, curvature, and homogeneity.

Properties of the profit function $\Pi(w,p)$:

1. $\Pi(w,p)$ is real-valued, continuous and differentiable with respect to w and p;

2. $\Pi(w,p)$ is non-increasing in w and non-decreasing in p for all $(w,p) > 0$. Thus, $\partial\Pi(w,p)/\partial p \geq 0$ and $\partial\Pi(w,p)/\partial w_i < 0$;

3. $\Pi(w,p)$ is linear homogeneous in (w,p). Thus, $\Pi(tw,tp) = t\Pi(w,p) \;\; t > 0$;

4. $\Pi(w,p)$ is convex in (w,p). Thus, the Hessian matrix of $\Pi(w,p)$ is positive semi-definite. Among other things this implies that $\partial^2\Pi(w,p)/\partial w_i^2 < 0 \quad \forall i$ and $\partial^2\Pi(w,p)/\partial p^2 > 0$.

Proofs:

Monotonicity. a) Let $w^1 \geq w^0$. We have: $\Pi(w,p) = \max\{py - w^1x\} \equiv py^* - w^1x^* = \Pi\left(w^1,p\right) \leq py^* - w^0x^* = \Pi\left(w^0,p\right)$; b) Let $p^1 \geq p^\circ$. We have: $\Pi(w,p) = \max\{p^1y - wx\} = p^1y^* - wx^* = \Pi\left(w,p^1\right) \geq p^0y^* - wx^* = \Pi\left(w,p^0\right)$. Note that these monotonicity properties follow immediately from invoking the envelope theorem for the maximum profit function. Indeed, note that: $\partial\Pi(w,p)/\partial p = y^* + \partial y/\partial p - \partial x/\partial p = y^* \geq 0$ as second order effects are negligible. Similarly, $\partial\Pi(w,p)/\partial p = \partial y/\partial w - x^* - \partial x/\partial w = -x^* \geq 0$.

Homogeneity. We multiply by any factor $t > 0$ output and input prices. We then compute the (maximum) profit function: $\Pi(t\cdot w, t\cdot p) = \max\{t\cdot py - t\cdot wx\} = t\cdot$

$\max\{py - wx\} = t \cdot \Pi(w, p)$. Recall from the profit maximization problem that the FOCs of the input choice are:

$$\begin{cases} \frac{\partial \pi}{\partial x_1} = f_1(x_1^*, x_2^*) - w_1/p = 0 \\ \frac{\partial \pi}{\partial x_2} = f_2(x_1^*, x_2^*) - w_2/p = 0 \end{cases} \tag{4}$$

If we multiply all prices by $t > 0$ we see that the profit maximizing input quantities do not change. It follows that via the production function also the profit maximizing output level does not change. Hence: $\Pi(tw, tp) = tpy^* - twx^* = t\Pi(w, p)$ which confirms the homogeneity property.

Convexity. Consider two sets of prices (p^0, w^0) and (p^1, w^1). For any $0 \le \lambda \le 1$ we have: $\Pi(\lambda w^0 + (1-\lambda)w^1, \lambda p^0 + (1-\lambda)p^1) \equiv \max\{[\lambda p^0 + (1-\lambda)p^1]\, y - [\lambda w^0 + (1-\lambda)w^1]x\} = [\lambda p^0 + (1-\lambda)p^1]y^* - [\lambda w^0 + (1-\lambda)w^1]x^* = \lambda(p^0 y^* - w^0 x^*) + (1-\lambda)(p^1 y^* - w^1 x^*) \le \lambda \Pi(p^0, w^0) + (1-\lambda)\Pi(p^1, w^1)$. The convexity of the profit function with respect to input and output prices is a consequence of the concavity of the cost function with respect to input prices and in turn of the substitutability of inputs. As illustrated in Figure 5.5, if the price of an output increases or the price of an input decreases and the firm does not change its production plan, profit will increase linearly. However, since the firm will want to re-optimize at the new prices, it can actually do better. Profits will increase at a greater than linear rate. Thus, the profit function is convex. Of course, recall that when input prices increase, maximum profits decrease.

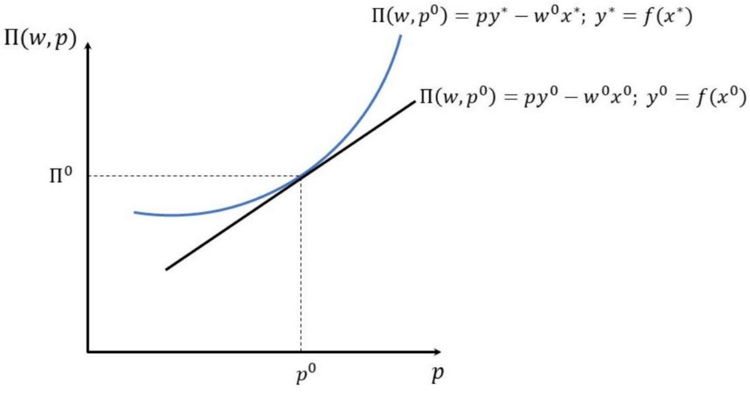

Figure 5.5 *Convexity of the profit function.*

5.6 Duality between production and profit functions

To prove the duality between production and profit functions we have to recall the properties of the technology. We can express these properties in terms of production possibility set Y or of transformation function. We will use this last representation. Thus:

Properties of the transformation function $T(y) = 0$, which we will denote synthetically $\boxed{\boldsymbol{T}}$:

1. $T(y)$ is a continuous function on $\{y : T(y) < 0\}$;
2. $T(y)$ is such that: (i) $T(y) = 0$ if y is on the frontier; (ii) $T(y) < 0$ if y is in the interior of Y; (iii) $T(y) > 0$ if y outside of Y;
3. $T(y)$ is convex (or at least quasi-convex).

These properties follow from Y being regular, monotonic and convex.

Properties of the profit function $\Pi(w, p)$, which we will denote synthetically $\boxed{\boldsymbol{\Pi}}$:

1. $\Pi(w, p)$ is real-valued, continuous and differentiable with respect to w and p;
2. $\Pi(w, p)$ is non-increasing in w and non-decreasing in p for all $(w, p) > 0$. Thus, $\partial\Pi(w, p)/\partial p \geq 0$ and $\partial\Pi(w, p)/\partial w_i < 0$;
3. $\Pi(w, p)$ is linear homogeneous in (w, p). Thus, $\Pi(tw, tp) = t\Pi(w, p) \quad t > 0$;
4. $\Pi(w, p)$ is convex in (w, p). Thus, the Hessian matrix of $\Pi(w, p)$ is positive semi-definite. Among other things this implies that $\partial^2\Pi(w, p)/\partial w_i^2 < 0 \quad \forall i$ and $\partial^2\Pi(w, p)/\partial p^2 > 0$.

We now have:

Theorem 1. If $T(y)$ satisfies Conditions $\boxed{\boldsymbol{T}}$, then the function $\Pi(w, p) = \max_{\{y,x\}} \left\{py - wx : (y, x) \in T(y) = 0\right\}$ satisfies Conditions $\boxed{\boldsymbol{\Pi}}$.

Theorem 2. If $\Pi(w, p)$ satisfies Conditions $\boxed{\boldsymbol{\Pi}}$, then the function $T(y)$ satisfies Conditions $\boxed{\boldsymbol{T}}$.

The theorems state that a profit function with the right properties corresponds uniquely to a well-behaved transformation function and in turn production set. These duality theorems together show that a profit function can completely describe a well-behaved technology. Indeed, a remarkable result in the theory of the firm is that if Y is convex, Y and $\Pi(w, p)$ contain the exact same information.

Let us illustrate the duality theorems for the case of a single input and a single output. Figure 5.6 shows the solution of the profit maximization problem at two different sets of prices $\left(p^0, w^0\right)$ and $\left(p^1, w^1\right)$. In the figure Y denotes the production set $Y = \left\{(y, x) \in R_+^2 | y \leq f(x)\right\}$. Our starting point, however, is that we know Π but we do not know f or Y. A procedure for finding the latter starting from the former consists of computing the profits associated to, say, $\left(p^0, w^0\right)$ equal to $\Pi\left(w^0, p^0\right) = p^0 y - w^0 x$ and represented by a line in Figure 5.6. All space above the line is eliminated because, if $\left\{(y, x) \in Y\right\}$ these input-output combinations cannot correspond to maximum profits at those prices. Repeat the same process for different prices $\left(p^1, w^1\right)$ and then again and again. We thus construct a well-behaved production set. Conversely, for every input level x the associated output level y can be on or below the frontier of Y. At any set of prices (w, p) the output levels outside the frontier of Y are not technologically feasible. Hence the input-output combinations on the frontier must generate the maximum profit for every set of prices.

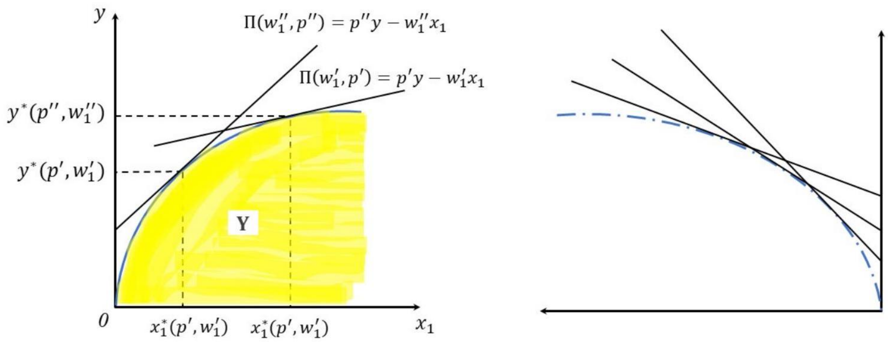

Figure 5.6 *Duality between well-behaved production sets and profit functions.*

5.7 Duality results using profit functions: Hotelling's lemma and properties of unconditional input demand functions

The main result of the duality between production and profit functions is known as Hotelling's lemma. It was originally formulated by Harold Hotelling in the paper "Edgeworth's taxation paradox and the nature of demand and supply functions" published in the Journal of Political Economy in 1932.

Hotelling's lemma enables to obtain the whole set of unconditional input demand functions and output supply functions by simple differentiation of the profit function with respect to input and output prices. That is:

Hotelling's lemma. If a profit function satisfies the conditions above and is differentiable with respect to all prices, then:

$$x_i(w, p) = -\frac{\partial \Pi(w, p)}{\partial w_i} \quad \forall i \tag{5}$$

$$y_j(w, p) = \frac{\partial \Pi(w, p)}{\partial p_j} \quad \forall j \tag{6}$$

where $x_i(w, p)$ is the unconditional demand functions for input i $(i = 1, \ldots, n)$ and $y_j(w, p)$ is the supply functions for output j $(j = 1, \ldots, m)$. Note that we have deliberately allowed for several outputs to emphasize the generality of the result.

Proof. As in the case of Shephard's lemma we can demonstrate the result by applying the envelope theorem. Alternatively, we can use the distance formula, which we do hereafter.

Define the distance function $g(p,w) = \Pi(p,w) - (py^* - wx^*)$ where (y^*, x^*) correspond to (p^*, w^*). Then $g(p,w) \geq 0$. At the maximum the necessary conditions require: $\frac{\partial g(p^*,w^*)}{\partial p_j} = \frac{\partial \Pi(p^*,w^*)}{\partial p_j} - y_j^* = 0$ and $\frac{\partial g(p^*,w^*)}{\partial w_i} = \frac{\partial \Pi(p^*,w^*)}{\partial w_i} + x_i^* = 0$. Note that the sufficient condition for a maximum requires the profit function to be convex in prices.

In summary, to derive the full system of output supply and factor demand functions simply specify a profit function, make sure it satisfies the regularity conditions $\boxed{\mathbf{\Pi}}$ and apply Hotelling's lemma. This procedure is crucially useful in applied work.

Properties of unconditional input demand functions $x(w,p)$ **and output supply functions** $y(w,p)$:

1. Unconditional input demand functions vary inversely with their own price: $\frac{\partial x_i(w,p)}{\partial w_i} \leq 0 \; \forall i$. This implies that input demand curves are downward sloped.

   Proof. By the convexity of the profit function and by Hotelling's lemma: $\frac{\partial x_i(w,p)}{\partial w_i} = \frac{-\partial^2 \Pi(w,y)}{\partial w_i^2} \leq 0$.

2. Output supply functions vary directly with their own price: $\frac{\partial y_j(w,p)}{\partial p_j} \geq 0 \; \forall j$. This implies that output supply curves are upward sloped.

   Proof. By the convexity of the profit function and by Hotelling's lemma: $\frac{\partial y_j(w,p)}{\partial p_j} = \frac{\partial^2 \Pi(w,y)}{\partial p_j^2} \leq 0$.

3. Cross price effects are symmetric: $\frac{\partial x_i(w,p)}{\partial w_h} = \frac{\partial x_h(w,p)}{\partial w_i}$, $\frac{\partial y_j(w,p)}{\partial p_s} = \frac{\partial y_s(w,p)}{\partial p_j}$ and $\frac{\partial x_i(w,p)}{\partial p_s} = \frac{\partial y_s(w,p)}{\partial w_i}$.

   Proof. By the symmetry of the cross partials of the profit function and by Shephard's lemma.

4. Unconditional input demand functions and output supply functions are homogeneous of degree zero in $(w,\ p)$, i.e.: $x(tw,tp) = x(w,y)$ and $y(tw,tp) = y(w,p)\quad t > 0$.

   Proof. By the properties of homogeneous functions, because the profit function is O^1 in (w,p), then its first partials (Hotelling's lemma) are O^0 in (w,p).

5.8 Additional duality results using profit functions

It is possible to show that there is an equivalent version of **Slutsky's decomposition** holding for profit and cost functions. Specifically:

$$\frac{\partial x^{\Pi}(w,p)}{\partial w_i} = \frac{\partial x^C(w,y)}{\partial w_i} + \frac{\partial x^C(w,y)}{\partial y}\frac{\partial y(w,p)}{\partial w_i} \quad \forall i \tag{7}$$

where the superscripts refer to either profit or cost functions. According to (7), the impact of a price change on the unconditional factor demand can be decomposed into the sum of the corresponding impact on the conditional factor demand and the product between the output effect on the conditional factor demand and the impact of the price change on the output supply.

It is possible to show that a crucial technological indicator can be recovered by an appropriate application of the profit function. Specifically:

$$\sigma_{ij} = \frac{\Pi_{ij}(p,w)\cdot\Pi(p,w)}{\Pi_i(p,w)\cdot\Pi_j(p,w)} = \frac{\Pi_{ij}(p,w)\cdot\Pi(p,w)}{x_i(p,w)\cdot x_j(p,w)} \quad \forall i,j \tag{8}$$

where the subscripts denote partial differentiation. The expression (8) represents the **partial elasticity of input substitution** if (i,j) both refer to inputs, the **partial elasticity of output transformation** if (i,j) both refer to outputs, and to the **partial elasticity of product intensity** if i denotes an input and j denotes an output.

We can obtain special cases of the profit function defined above. Specifically, let (all) output levels be given $y = \overline{y}$. Then, the profit function:

$$\begin{aligned}\Pi(w,p) &= \max_{\{y,x\}}\left\{p\overline{y} - wx : \left(\overline{y},x\right) \in T(y) = 0\right\}\\ &= \max_{\{x\}}\left\{-wx : \left(\overline{y},x\right) \in T(y) = 0\right\}\\ &= \min_{\{x\}}\left\{wx : \left(\overline{y},x\right) \in T(y) = 0\right\}\\ &= C(w,\overline{y})\end{aligned}$$

which is the minimum cost function. Similarly, let (all) input levels be given $x = \overline{x}$. Then, the profit function:

$$\begin{aligned}\Pi(w,p) &= \max_{\{y,x\}}\left\{py - w\overline{x} : \left(\overline{y},x\right) \in T(y) = 0\right\}\\ &= \max_{\{x\}}\left\{py : \left(\overline{y},x\right) \in T(y) = 0\right\}\\ &= R(p,\overline{x})\end{aligned}$$

This is referred to as the maximum **revenue function**. One can find (few) empirical applications of it especially in the agricultural economics literature, where the fixed factor is usually land.

Finally, partition the vector of input quantities as $x = \left[x_v, x_f\right]$ and correspondingly $w = \left[w_v, w_f\right]$. The distinction is between input levels that can be changed immediately at no additional cost and input levels that can be only eventually changed. This distinction is familiar when distinguishing between short run where $x_z = \overline{x}_z$ and the long run where there is no longer any distinction/partition and all inputs can be freely change, i.e. are all variable inputs.

On the basis of the above constraints on input and output levels we can further define the **restricted or short run cost function**:

$$C\left(w_v, y, x_z\right) = \min_{x_v}\left\{ww_v : \left(\overline{y}, x_v, \overline{x}_f\right) \in T(y) = 0\right\}$$

and the **restricted or short run profit function**:

$$\Pi\left(w_v, p, x_z\right) = \max_{\{y, x_v\}} \left\{py - ww_v : \left(y, x_v, \overline{x}_f\right) \in T(y) = 0\right\}$$

We can reformulate the profit function and the duality with the production set by considering **netputs** y, which can be either positive (outputs) or negative (inputs). The corresponding prices are denoted by p.

The profit function $\Pi(p)$ is the solution of the following profit maximization problem $\Pi(p) = \max_y \left\{py | y \in Y\right\} = p \cdot y(p)$, where $y(p)$ is called **net supply function**.

The **profit function $\Pi\left(p\right)$ has the following properties**:

1. $\Pi(p)$ is increasing in p which means increasing in output prices and decreasing in input prices;

2. $\Pi(p)$ is homogenous of degree 1 in p. Since scaling all prices by $t > 0$ does not change relative prices, the optimal production plan does not change. However, multiplying all prices by t multiplies profit by t as well. Hence the linear homogeneity;

3. $\Pi(p)$ is convex in p. The reason is that, if the price of an output increases (or an input decreases) and the firm does not change its production plan, profits will increase linearly. However, since the firm will want to re-optimize at the new prices, it can actually do better. Profits will increase at a greater than linear rate.

The **net supply function $y(p)$ has the following properties**:

1. $y(p)$ is homogeneous of degree zero in p. The reason for this is the same as always: the optimality conditions for the profit maximization problem involve a tangency condition, and tangencies are not affected by re-scaling all prices by the same amount;

2. Hotelling's lemma: $y_j(p) = \frac{\partial \Pi(p)}{\partial p_j} \quad \forall j$. The lemma relates the profit function $\Pi(p)$ to the net supply functions $y(p)$. The proof follows directly from the envelope theorem;

3. If $y(p)$ is single-valued and differentiable, the matrix of second derivatives $\Pi(p)$, with typical term $\partial^2 \Pi(p)/\partial p_j \partial p_h$, is a symmetric, positive semi-definite matrix. This follows from the convexity of $\Pi(p)$ in prices. Note that this implies that the diagonal elements of the matrix $\partial y_i(p)/\partial p_i \geq 0$, i.e. net supply functions increase with output prices and decrease with input prices;

4. If $y(p)$ is single-valued and differentiable, then $\partial y_i(p)/\partial p_h = \partial y_h(p)/\partial p_i$, i.e. cross prices effects of the net supply functions are symmetric. This follows from Hotelling's lemma and from the symmetry of the matrix of second derivatives $\Pi(p)$.

5.9 Duality theory of producer behavior: summary

To summarize, we visualize the duality existing between production sets and cost functions on the one hand and with profit functions on the other hand. These two functions are in turn related as shown in Figure 5.7.

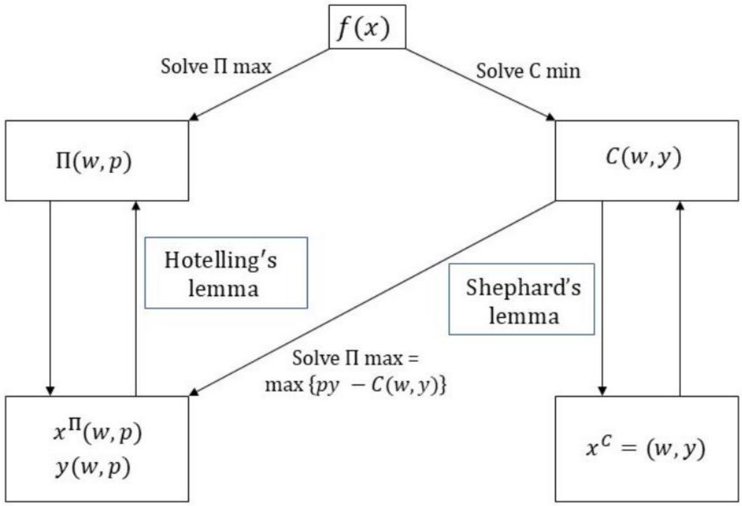

Figure 5.7 *Duality in production theory.*

5.10 Using dual cost functions to characterize a firm's technology

We noted before that dual cost functions incorporate information on the features of the firm's technology. A powerful implication is that we can infer from the study of the cost function features of the technology such as the degree of input substitution, the degree of scale returns, and the potential homotheticity or homogeneity of the production function, among others. The cost function approach actually turns out to be very useful for empirical purposes. We often have better data on the costs a firm incurs than on its production function. But, thanks to duality results, we know that the cost information contains everything we want to know about the firm's technology. In addition, recall that the profit maximization is not very useful for technologies with increasing returns to scale. This is not so with the cost minimization problem, which is perfectly well-defined as long as the production function is quasi-concave, even if it exhibits increasing returns to scale. Hence the cost function may allow us to say things about firms when profit maximization does not. The cost approach is useful for another reason. When the output price p is fixed, the firm's profit maximization problem is: $\Pi = py - C(w, y)$. However, this approach can also be used when the price the firm charges is not fixed. In particular, suppose the price the firm can charge depends on the quantity it sells according to some function $p(y)$. The profit maximization problem can be written as: $\Pi = p(y)y - C(w,\ y)$. Thus, our study of the cost function can also be used when the firm is not competitive, as in the cases of monopoly and oligopoly. It remains true that one limitation of the cost approach is that output is taken as exogenously determined, which is not realistic. In addition, the firm always takes input prices as given.

We now use the cost function and its properties to characterize many of the features of a firm's technology. First of all, Figure 5.8 reminds us that the cost minimization problem has a dual conjugate problem consisting of looking for the maximum output obtainable at given cost of the factors involved.

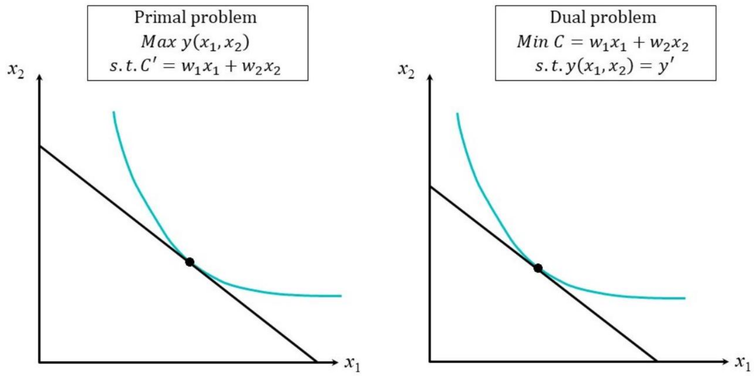

Figure 5.8 *Cost minimization and output maximization.*

1. Homothetic and homogenous technologies

We begin with the homotheticity and the homogeneity of production functions. We already defined homothetic utility functions.

A production function $F(x)$ is said to be homothetic if it can be written $F(x) = \varphi\left[f(x)\right]$ where $f(x)$ is homogenous of degree 1 and $\varphi[.]$ is a continuous positive monotonically increasing function of $f(.)$. Then we have the following.

Result 1. If the production technology $F(x)$ is homothetic, then the dual cost function can be written as $C(w, y) = \varphi^{-1}(y)\cdot C(w,\ 1) = \varphi^{-1}(y)\cdot c(w)$, where $c(w)$ satisfies the properties for a well-behaved (unit) cost function. Thus, cost functions of homothetic technologies can be written as the product of two functions, one depending only on output levels and the other only depending on factor prices.

Proof. The cost function is defined as the solution of the problem:

$$\begin{aligned}
C(w, y) &= \min_x \left\{wx : \varphi\left[f(x)\right] \geq y\right\} \\
&= \min_x \left\{wx : f(x) \geq \varphi^{-1}(y)\right\} \\
&\text{multiply and divide by } \varphi^{-1}(y): \\
&= \min_x \left\{\varphi^{-1}(y)\, w\left[x/\varphi^{-1}(y)\right] : f(x)/\varphi^{-1}(y) \geq 1\right\} \\
&= \varphi^{-1}(y) \min_x \left\{w\left[x/\varphi^{-1}(y)\right] : f(x)/\varphi^{-1}(y) \geq 1\right\} \\
&= \varphi^{-1}(y) \min_x \left\{wz : f(z) \geq 1\right\}
\end{aligned}$$

where $z = x/\varphi^{-1}(y)$:

$$= \varphi^{-1}(y)C(w,1)$$

Result 2. If the production technology $F(x)$ is linear homogenous, i.e. it exhibits constant returns to scale, then the cost function dual to a constant returns technology can be written as $C(w,y) = y \cdot C(w,1) = y \cdot c(w)$.

Proof. Recall from the definition of homothetic technologies that $F(x) = \varphi\left[f(x)\right]$ where $f(x)$ is homogenous of degree 1. Now let $\varphi\left[f(x)\right] = f(x)$, i.e. $\varphi[.] = 1$. Hence the result.

Result 3. If the production technology $F(x)$ is homogenous of degree k, then the dual cost functions of homogenous technologies can be written as $C(w,y) = y^{1/k} \cdot C(w,1) = y^{1/k} \cdot c(w)$.

Proof. The cost function is defined as the solution of the problem:

$$C(w,y) = \min_x \left\{wx : F(x) \geq y\right\} = \min_x \left\{wx : y^{-1}F(x) \geq 1\right\}$$

Homogeneity of degree k implies that $f(tx) = t^k f(x)$. Let $t = y^{-1}$. Thus:

$$= \min_x \left\{y^{1/k} \cdot wz : f(z) \geq 1\right\}$$

where $z = y^{-1/k}x$. Thus:

$$= y^{\frac{1}{k}} \cdot C(w,1) = y^{\frac{1}{k}} \cdot c(w)$$

The degree of returns to scale of a technology is therefore given by the reciprocal of the exponent of output of the above cost function.

2. Returns to scale

We now consider two general features of any production technology: the local returns to scale and the degree of input substitutability.

Result 4. The proper measure of returns to scale at a point is the reciprocal of the elasticity of costs with respect to output. Specifically, the average cost function is increasing (resp. decreasing) in y iff (if and only if) the elasticity of costs with respect to output is less (resp. greater) than unity.

Proof. Compute:

$$\frac{\partial}{\partial y}\frac{C(w,y)}{y} = \frac{y \cdot \frac{\partial C(w,y)}{\partial y} - C(w,y)}{y^2}$$

$$= \frac{C(w,y)}{y^2} \cdot \left[\eta(w,y) - 1\right]$$

where:

$$\eta(w,y) = \frac{\partial C(w,y)}{\partial y} / \frac{C(w,y)}{y}$$

$\eta(w,y)$ is the cost elasticity of output, given by the ratio between marginal and average costs. Thus, an elasticity greater (resp. smaller) than one denotes local decreasing (resp. increasing) returns to scale.

It should be noted that in general the cost elasticity provides a local measure as it depends on output levels and factor prices. However:

Result 5. If $F(x)$ is homothetic, then $\eta(w,y)$ depends only on y, not on w.

Proof. Recall that if $F(x)$ is homothetic, from Result 1 above $C(w,y) = \psi(y) \cdot c(w)$. It follows that $\eta(y) = y \cdot \frac{\partial \psi(y)}{\partial y} / \psi(y)$ which is independent of w.

Result 6. If $F(x)$ is homogenous of degree k, then $\eta(w,y) = k^{-1}$, a constant independent of both y and w.

Proof. From Result 2 above $C(w,y) = y^{1/k} \cdot c(w)$. It follows that $\eta(y) = \left(\frac{1}{k}\right) y \cdot y^{(1/k)-1} / y^{1/k}$ which is independent of w.

3. Elasticity of input substitution

We now turn to input substitution. Recall that, in the case of n inputs the Allen-Uzawa partial elasticity of substitution (AUES) between inputs i and j is defined as:

$$\sigma_{ij} = \frac{\sum_{h=1}^{n} F_h x_h}{x_i x_j} B_{ij}^{-1} \tag{9}$$

where $F_h = \partial F(x)/\partial x_h$ is the marginal product of input h, $F(x)$ is the usual production function and B is the Bordered hessian of $F(x)$:

$$B = \begin{bmatrix} 0 & f_1 & \cdots & f_n \\ f_1 & f_{11} & & f_{1n} \\ \vdots & & \ddots & \vdots \\ f_n & f_{n1} & \cdots & f_{nn} \end{bmatrix}$$

Result 7. The AUES can be expressed using cost functions as:

$$\sigma_{ij}^{C} = \frac{C_{ij}(w,y) \cdot C(w,y)}{C_i(w,y) \cdot C_i(w,y)} \tag{8}$$

where $C_{ij}(w,y) = \partial^2 C(w,y)/\partial w_i \partial w_j$ and $C_i(w,y) = \partial C(w,y)/\partial w_i$.

Proof. Consider the FOCs of the cost minimization problem:

$$w_i - \lambda F_i(x) = 0$$

$$y - F(x) = 0$$

Totally differentiating the FOCs yields:

$$\lambda \begin{bmatrix} 0 & f_1 & \cdots & f_n \\ f_1 & f_{11} & & f_{1n} \\ \vdots & & \ddots & \vdots \\ f_n & f_{n1} & \cdots & f_{nn} \end{bmatrix} \begin{bmatrix} d\lambda/\lambda \\ dx_1 \\ \vdots \\ dx_n \end{bmatrix} = \begin{bmatrix} \lambda dy \\ dw_1 \\ \vdots \\ dw_n \end{bmatrix}$$

Solving:

$$\begin{bmatrix} d\lambda/\lambda \\ dx_1 \\ \vdots \\ dx_n \end{bmatrix} = \frac{1}{\lambda} B_{ij}^{-1} \begin{bmatrix} \lambda dy \\ dw_1 \\ \vdots \\ dw_n \end{bmatrix}$$

This yields:

$$\frac{\partial x_i}{\partial w_j} = \frac{1}{\lambda} B_{ij}^{-1}$$

Using the definition of AUES σ_{ij} and substituing λ from thee FOCs and from the input price effect above, we can write:

$$\sigma_{ij} = \frac{\sum_{h=1}^{n} w_h x_h}{x_i x_j} \frac{\partial x_i}{\partial w_j}$$

Applying Shephard's lemma we obtain (8).

Result 8. The **price elasticity of the demand for input** i with respect to price j can be expressed as follows:

$$\epsilon_{ij}(w, y) = s_j \cdot \sigma_{ij}(w, y) \text{ where } s_j = w_j x_j / C(w, y)$$

Proof. Multiply and divide $\sigma_{ij} = \frac{C(w,y)}{x_i x_j} \frac{\partial x_i}{\partial w_j}$ by $\left(\frac{x_i}{w_j}\right)$. Then:

$$\sigma_{ij} = \frac{C(w, y)}{x_i x_j} \frac{\partial x_i}{\partial w_j} \frac{w_i}{x_j} \frac{x_i}{w_j} = s_j^{-1} \epsilon_{ij}(w, y)$$

Result 9. If $C(w, y)$ is obtained from a **homothetic technology**, then $\sigma_{ij}(w)$ is independent of y.

Proof. From Result 4 above we have $C(w, y) = \psi(y) \cdot c(w)$. Hence:

$$\sigma_{ij} = \frac{c_{ij}(w) \cdot c(w)}{c_i(w) \cdot c_j(w)}$$

4. Separability

We previously defined weakly and strongly separable technologies. This is an issue with potential empirical relevance. We ask here if it is valid, i.e. consistent with the underlying technology, to group input levels or input prices in production or cost functions. Consider the following cases:

Case 1. Suppose $y = f(L_U, L_S, K)$ where L_U and L_S are unskilled and skilled labor inputs respectively. Can we represent this technology with the production function $y = f[L(L_U, L_S), K] = f(\overline{L}, K)$ where $\overline{L} = L(L_U, L_S)$ is an aggregate labor input?

Case 2. Consider the so-called KLEM model $Y = F(K, L, E, M)$ where Y is gross output, K is fixed capital, L is labor, E is energy, and M is materials. We ask: a) Is it valid to write this as a function of an "aggregate input" value added, energy, and materials. i.e. $Y = F[f(K, L), E, M] = F(VA, E, M)$, so that we can study the substitution between K and L without reference to E or M? b) If it is valid, can we similarly write the dual cost function as $C = h(w_K, w_L, w_E, w_M, y) = h[v(w_K, w_L), w_E, w_M, y]$?

Case 3. Consider the KLEM-type model where $Y = F(K, L, e_1, e_2, \ldots, e_s, M)$ and the e_i ($i = 1, .., s$) are individual energy sources (natural gas, oil, coal, electricity, etc.). We ask: a) Is it valid to write this technology as a function of an "aggregate energy input", i.e. $Y = F(K, L, E, M)$ where $E = E(e_1, e_2, \ldots, e_s)$, so that we can study the substitution between individual energy sources without reference to K, L or M?

Recall the definitions of functional separability of production technologies given before.

Weak separability. Given $y = F(x)$ suppose we can group the elements of the input vector x into N groups: $x^1, x^2, \ldots, x^N$ such that $y = F[g^1(x^1), g^2(x^2), \ldots, g^N(x^N)]$. If $F(.)$ Is **homothetic** we say that $F(.)$ is **weakly separable** into the N partitions of x. If instead $F(.)$ Is **non-homothetic** we say that $F(.)$ is **homogenously separable** into the N partitions of x.

Result 10. If $F(.)$ is weakly separable then the **dual cost function** can be written as $C(w^1, w^2, \ldots, w^N, y) = \psi(y) c\left[h^1(w^1), h^2(w^2), \ldots, h^N(w^N)\right]$. If $F(.)$ is homogeneously separable then the **dual cost function** can be written as $C(w^1, w^2, \ldots, w^N, y)$ $= C\left[h^1(w^1), h^2(w^2), \ldots, h^N(w^N), y\right]$.

Strong separability. A production function $y = F(x)$ is strongly separable if $F(.)$ is **homothetic** and can be written $y = F[\sum_{k=1}^N g^k(x^k)]$ where $g^k(.)$ are regular production functions.

Result 11. If $F(.)$ is strongly separable then the **dual cost function** takes the form $C(w^1, w^2, \ldots, w^N, y) = \psi(y) c[\sum_{k=1}^N g^k(w^k)]$.

Bibliographic notes

In these lecture notes there are no references except for two popular Microeconomics textbooks, one at intermediate level and the other at advanced level. They are from the same author:

- Varian, Hal R. *Intermediate Microeconomics. A Modern Approach* (ninth edition). W. W. Norton & Company, 2014.
- Varian, Hal R. *Microeconomic Analysis* (third edition). W. W. Norton & Company, 2009.

The latter is the standard U.S. graduate level Microeconomics textbook usually adopted in alternative to:

- Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green. *Microeconomic Theory*. Oxford University Press, 1995.

There are several other advanced textbooks that deal with the material of these notes, more or less extensively. Among those that present also duality theory I would mention two books:

- Jehle, Geoffrey A., and Philip J. Reny. *Advanced Microeconomic Theory* (third edition). Prentice-Hall, 2011.
- Cowell, Frank A. *Microeconomics. Principles and Analysis* (second edition). Oxford University Press 2018.

The literature on consumer and producer behavior is vast. There are many journal articles, survey papers, edited volumes, monographs. I will stay away from any attempt to list some of them, with only three exceptions that I think are useful guides:

- Chambers, Robert G. *Applied Production Analysis: A Dual Approach*. Cambridge University Press, 1988.
- Cornes, Richard. *Duality and Modern Economics*. Cambridge University Press, 1992.
- Deaton, Angus, and John Muellbauer. *Economics and Consumer Behavior*. Cambridge University Press, 2012.

Lecture Notes on the Theory of Consumer and Producer Behavior

Marzio Galeotti

These notes present a fairly advanced treatment of the microeconomic theory of consumer and producer behavior. They are meant to satisfy two goals: on the one hand they could provide support material to Ph.D. students in Economics tackling their Microeconomics sequence, on the other hand they could provide a tool for researchers in Economics who would like to deepen their knowledge of the theory of consumer and producer decisions. One special feature of these notes is that they include both the standard analysis of consumer and producer behavior – as found in advanced microeconomics textbooks – and the so-called duality approach, which proves especially useful in applied and empirical work.

ISBN 979-12-5510-395-0 (print)
ISBN 979-12-5510-389-9 (PDF)
ISBN 979-12-5510-393-6 (EPUB)
DOI 10.54103/milanoup. 286