PART I: Neoclassical Price Theory

I. Introduction: A Short History of Price Theory

Attempting to understand market prices is an ancient endeavor—as old as markets. Every seller of goods and services has an interest in learning about the highest price at which their products can be sold and the lowest at which they can be sold without losing money. Thus, for thousands of years before general economic models were worked out in the nineteenth century, there were intuitive models worked out by individual proprietors and entrepreneurs for their own products and markets. They all realized that the prices at which they could sell their goods would be higher when buyers had relatively intense desires for their goods and lower when they had to compete with other craftsmen and merchants for the purchases of potential buyers.

Business necessities and common sense also would have linked their minimum selling prices to their cost of production. To be profitable, their selling price had to be greater than a good's average cost, otherwise there would be no net revenue associated with bringing products to market. Ideally—from the perspective of sellers—the selling price would be much higher than their production costs; however, they also realized that setting prices "too high" would reduce their sales to a pittance or to zero. Thus, arithmetic and common sense implied that pricing was in practice constrained to a relatively narrow band of "reasonable" or "fair" prices by the limitations of consumer demand, the cost of producing goods and services, and competition from rival craftsmen and merchants.

After general pricing or markup rules of thumb for particular products were developed, they would be passed on to successive generations of artisans, merchants, and middlemen. These rules of thumb provided the first economic theories of market prices.

More general principles for understanding market prices were subsequently developed by entrepreneurs with a scholarly bent and by scholars interested in how markets operated. Among the latter, early and relatively well-known examples are found in Aristotle's books on politics and ethics. Aristotle's writings provide a theory of money goods and inflation and also a theory of relative prices that is worked out to illustrate his concept of relative justice. In his theory, the band of "fair" relative prices was a narrow one—essentially that associated with zero profits for speculators and middlemen. His brief discussion of relative prices is largely compatible with contemporary theories of general equilibria in perfectly competitive markets. Differences in market structures were also understood. For example, in a footnote, Aristotle mentions a fellow philosopher who monopolized the market for olive presses, so that he could demonstrate that philosophers could be rich if they wanted to be. That footnote demonstrates that a theory of monopoly pricing also existed at that time. Monopoly prices tend to be significantly higher than those in otherwise similar competitive markets.

Athens and several other Greek city states were centers of commerce during Aristotle's lifetime, with trading posts distributed around the shores of the Mediterranean Sea. Athens was also a center of intellectuals, schools, and scholarship. Thus, it was natural that at least a few Greek scholars were interested in the general properties of market networks—even when it was not a main focus of their writings or speculations. Aristotle claimed no originality in his short sections on economics and mentions one or two scholars who were more interested in the behavior of markets than he was.

Contemporary efforts to develop general principles that account for market prices originated in the eighteenth century. These efforts arguably began with Adam Smith's most famous book, which was published in 1776, *the Wealth of Nations*. He was not the first or only thoughtful person interested in markets, but he is the first to write a long book on the subject complete with theories (general principles and models) and empirical support for his theory. In that book, Adam Smith develops a theory of long-run price determination based on scarcity and production costs. The more hours of work required to produce something, the more it would cost relative to products that required fewer hours to assemble. He also noted that the number of hours required varied with technology and specialization. He also developed a theory of the effect of money goods on long term prices and price fluctuations.

It should be kept in mind that during Smith's time, most "skills" could be learned through relatively short periods of apprenticeship, and thus differences among types of labor could be neglected without much loss. Commerce was growing in northwestern Europe, but most people were still farmers, farm workers, or servants in that period. Specialized firms and commercial centers had, nonetheless, long existed, and their numbers and sizes were expanding in England during his lifetime.

It should also be kept in mind that Smith was not attempting to provide a model that would precisely describe prices—possibly because he did not believe that one was possible. His aim was simply to characterize equilibrium prices in open competitive markets in the long run. He concludes that competition in the English context provided an invisible hand that guided the economy towards prosperity, because it was not unduly restrained by governmental regulations.

The Marginal Revolution

Economics, to the extent it was studied in the eighteenth and nineteenth centuries, was undertaken by philosophers such as Adam Smith and Jeremy Bentham, clergymen such as Thomas Malthus, practical businessmen such as David Ricardo, and successful bureaucrats such as John Stuart Mill, all of whom had sufficient time and interest in markets to speculate a bit about how they operated. It was a field of study, but not one favored by many academics or other intellectuals.

One field of philosophy was especially influential in the emergence of economics as a distinct field of study during the nineteenth century.

The concept of utility was at the heart of utilitarianism, an important strand of normative theory promoted by Jeremy Bentham, John Stuart Mill, and many others during the nineteenth century. It argued that all humans attempt to maximize utility—a single overarching index of all human aims and purposes—and that the best society was the one that produced the greatest sum of utility for its residents (or the greatest average happiness). Several of the leading utilitarian theorists wrote economic textbooks including Bentham (1793) and Mill (1848). The early utilitarians understood that utility was the basis of demand and mutual gains from trade, but not that marginal utility played a central role in the determination of prices. The "water-diamond" paradox troubled them, although it did not undermine their basic ideas about the foundations of demand, the voluntary nature of exchange, and their effects on aggregate utility (gains from trade necessarily increase it).

About a century after Adam Smith's treatise was published, three scholars independently worked out economic theories based on the logic of diminishing marginal utility: William Jevons (1871), Carl Menger (1871) and Léon Walras (1874).

In Jevons's case, he worked out the mathematics behind utility maximization which led directly to the mathematical notion of marginal utility—as the first derivative of a utility function with respect to a good or service. It turned out that marginal utility, according to Jevons, largely determined an individual's decision about the quantities of goods and services that individuals will purchase. Menger (1871) emphasized the subjective nature of both utility and marginal utility. Walras (1874) subsequently extended the idea to develop a theory of general equilibrium for an entire marketplace, with prices simultaneously clearing all markets. Prices, he argued, would adjust through trial and error (e. g. through a *tâtonnement* process).

It can be argued that after these three contributions by the early marginalists and the earlier ones by classical economists, the field of microeconomics and many of its core results had been established—as evident in Alfred Marshall's widely used textbook.

However, it took another 75 years for the neoclassical synthesis to be generalized and be fully worked out. Indeed, it is still being worked out.

II. Organization and Focus of Part I

Part I of this text reviews the core ideas that emerged from the neoclassical synthesis. Most of the models developed were refinements of the work of Jevons, Walras, and Marshall that were worked out during the first half of the twentieth century. It was also in this period that contemporary microeconomics textbooks came to have their present focus and used a combination of geometry and calculus to demonstrate important economic relationships. This mathematization of its core models caused economic theory to resemble a "hard science" like physics, although its models of volition yielded predictions that were rarely as precise as the best models of inanimate phenomena developed by physicists. (This resemblance is partly because Paul Samuelson, one of the pioneers of the neoclassical synthesis, was trained as a physicist.)

Part I consists of 4 chapters. Chapter 2 develops the neoclassical theory of demand. Chapter 3 develops the neoclassical theory of supply, and Chapter 4, develops the neoclassical theory of production. Chapter 5 provides an overview of their implications for equilibrium prices. The models reviewed in these chapters are of the partial equilibrium variety, largely because these are the easiest to understand and model. However, the appendix to Chapter 5 provides a short illustration of the logic and mathematics of general equilibrium analysis. Each chapter in Part I begins with geometric illustrations, then develops mathematical results initially using concrete functional forms, followed by models grounded in more abstract families of functions. The latter provide the basis for the most general results.

Although mathematics played (and continues to play) an important role in the development of price theory, readers should focus at least as much attention on the logic of human behavior and market outcomes being characterized. Readers should also be alert for the implicit assumptions used to produce the models developed—not all of which are completely obvious. For example, the optimization problems that undergird the analyses implicitly assume that all market participants are well-informed. They know their objective functions (utility and profits) and constraints (budgets and technology). There are no uncertainties or random aspects of their choice settings. Softening or dropping these assumptions provides part of the motivation for the models developed in parts II and III of the textbook, most of which were worked out in the second half of the twentieth century.

Chapter 2: Net Benefits, Utility, and Market Demand

I. Introduction: Rational Choices

At the heart of neoclassical economics is the "rational man," an asexual personage that makes decisions by considering the consequences of his or her action using an internally consistent yardstick to gauge the relative merits of those consequences. That yardstick is usually said to be calibrated in "utils" rather than inches or centimeters, which reflects the utilitarian origins of much of neoclassical economics. A rational "man" always chooses the best action—as judged by his or her internal personal utility yardstick. Rational individuals are, thus, said to be utility maximizers. Their choices all maximize utility (the unified goal of a human's action), given the constraints associated with the choice setting in which decisions are made and actions taken.

For the most part, the neoclassical theory of consumption assumes that consumers are well-informed. Thus, the prices of all relevant products are known, and the utility generated by successive units of each relevant product is also well-understood. In markets for routine purchases, this is an entirely reasonable assumption. In such cases, the consequences of purchasing and using each product acquired are exactly what he or she expects.

In extensions of the core model, a consumer might have to take account of unobservable random characteristics of the things bought that affect the quality of the good or service acquired (its marginal utility). An apple, for example, may be more or less ripe or more or less tasty in a way that is not observable until one bites into it. The durability of a car or laptop may not be known until it has been used for many years. Or, they might consider purchasing unfamiliar products (as when one travels to another country). Is a new cell phone really better than my old one? Is AI product A better than AI product G, and are such products really worth their cost?

In such cases, individuals may attempt to maximize expected or average utility rather than realized utility, because they cannot be certain about the quality (marginal utility) of the products on offer. Such cases are analyzed in Part II of the book. In those cases, individuals can be rational in two senses. They make internally consistent choices (are rational in the first sense) and, they have unbiased expectations (rational in a second sense). Or, they may be assumed to make the best use of available information about the consequences of purchasing and using goods and services—which when incomplete may lead to errors and regrets. This book uses the term rationality mainly in the first sense—although reasonable expectations (not too biased or error prone) help ensure that rationality in the first sense is observed.

In most of the models developed in the twentieth century, consumers are assumed to have "perfect" information. They know as much as can be known at the time that the actions of interest are taken. They may know exactly what they are purchasing or know the probability distribution of all the factors that affect its marginal utility. Other cases, as when there are unknown unknowns, are not entirely ignored, but neoclassical models nearly always begin by examining settings in which individuals are perfectly informed. And, it should be acknowledged that textbooks often end where they begin—e.g., with the perfect information case. This textbook will analyze some of the other cases in part II.

As long as the products being purchased are familiar ones, "perfect information" is a reasonable assumption, and it is the one used throughout Part I. The economic effects of imperfect information are taken up in Part II.

II. The Geometry of Net Benefit Maximization and Consumer Demand

The geometric approach to economics began in the late nineteenth century and continues to be used in undergraduate textbooks, although it is rarely used in articles published in economic journals today. When diagrams are included in major economic journals, they are used to illustrate what is being derived using other mathematical tools. Such illustrations are useful because most people's geometric intuition is better than their calculus, differential equation, or real analysis intuitions.

Thus, the chapters of Part I all begin with geometric derivations and diagrams. This approach also helps those whose undergraduate economic courses relied on geometry to make the transition from geometric to calculus-based economic models.

In general, geometric derivations assume that decision makers are either "net benefit" maximizers or "utility" maximizers. A person that maximizes net benefits maximizes the difference between their benefits measured in some currency, as with dollars, euros, yen, or yuan, etc. and its costs measured in the same currency. Although the geometric approach can deal with discrete units as well as infinitely divisible units of goods and services, by using lines to characterize marginal costs and marginal benefits, geometric models implicitly assume that goods and services are infinitely divisible. This allows the geometric results to serve as a preview of many of the mathematical results that follow them in Part I, which make the same implicit assumption. (Line-based diagrams are also, of course, the most common form depicted in undergraduate textbooks.) The assumption of "infinite divisibility" usually does not affect the results very much—although occasionally it does—but it does allow a variety of mathematical methods to be used in the modeling process—ones that most students will have learned in high school or as undergraduates in college.

In practice, goods tend to be sold in prepackaged discrete units or sizes, although it is still possible to purchase fruit and in some types of hardware by the pound or kilogram. Thus, there are still goods sold in any amounts that can be measured. It is this latter case that most mathematical models assume, partly because it makes it possible to use calculus to build models as demonstrated throughout the book. The discreteness of many products and inputs can be neglected in most cases, without generating significant errors, because the logic of diminishing marginal utility and diminishing marginal product applies to both the discrete and continuous cases.

The term "marginal" is an adjective that means the change in something—such as benefit, cost, utility, output—associated with a one unit change in quantity purchased or used as an input in cases in which goods are sold only in discrete units. In the continuous case, it is the rate at which that variable (benefit, cost, utility, etc.) changes as quantity purchased, produced, or used as an input increases.

Maximizing Net Benefits

In the first models of consumer choice developed in this chapter, each quantity that might be purchased by a consumer is assumed to have a particular (total) benefit level associated with it and a particular (total) cost associated with it. Both benefits and costs are assumed to be measured in dollars, euros, yen, or another currency. This allows one to represent the benefits and costs associated with the quantity of a particular good with two different functions, one that maps quantities into benefits, and the other that maps quantities into costs. The first is referred to as the total benefit or total value function and the second as the total cost function. The difference between total benefits and total costs is an individual's net benefit.

The total benefit function is assumed to be strictly concave. **A strictly concave function has the property that** a line (cord) connecting any two points on such functions always lies below the function. Strict concavity is an implication of both diminishing marginal utility and diminishing marginal benefit. A consumer's total cost function is normally assumed to be either linear (a straight line) or **convex**. A line connecting any two points on the graph of a convex function will lie on or above the function. Convex cost functions are an implication of diminishing returns in production, as developed in Chapter 4.

The intuition behind diminishing marginal utility is that an individual will use the first unit of a good for the purpose that generates the highest benefit, the second unit for the application with the second highest benefit, the third for that with the third highest benefit, and so on. Differentiability is not necessary for individuals to behave as net-benefit maximizers; it is simply mathematically convenient to assume that the total benefit and total cost functions are differentiable. It allows ideas from geometry and calculus to be used to analyze the decisions of such consumers.[1]

The marginal cost of a good is the price paid for successive units of that good. The price may be a subjective one—the opportunity cost of spending additional time, attention, and/or money to acquire the good—or it may be an objective measure—the amount paid for successive units of the good. Initially, we'll use the simpler objective measure, but later in the chapter we'll show that the opportunity cost measure is in some ways more general and useful.

Figure 2.1: Maximizing Net Benefits

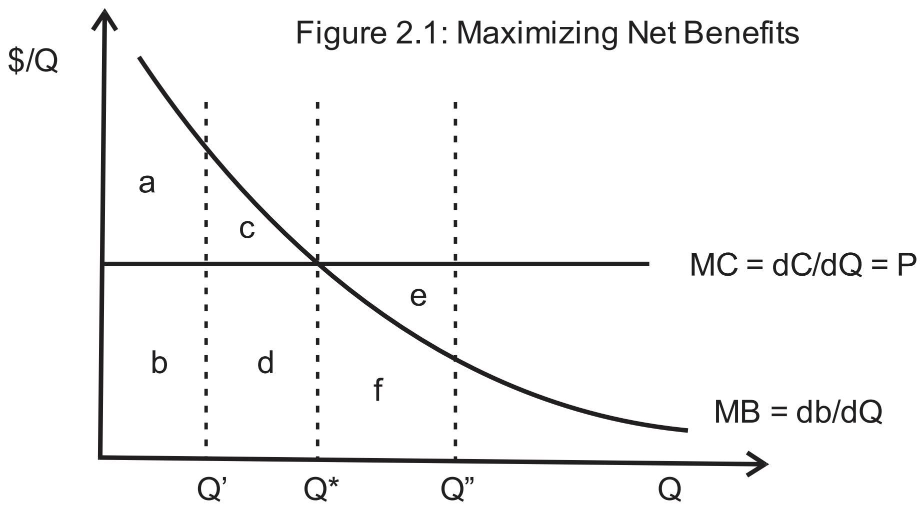


Figure 2.1 illustrates a typical marginal benefit curve and marginal cost curve for a representative individual, who will be called Al (short for Allan or Alice). It turns out that the area under the marginal benefit curve from 0 to any quantity Q is the total benefit of Q units of the good or service of interest [b(Q)] and that the area under the marginal cost curve is the total cost of Q units of the good [c(Q)]. Subtracting total cost from total benefit is

[1] For example, it assumes that one unit of a good can be applied to any use. However, if a particular number of the good completes a set and therefore allows a previously impossible use to be realized, it is possible that the units that complete a set have a higher marginal utility and marginal benefit than the previous unit(s).

Al's net benefit from purchasing particular quantities of the good or service of interest. In a diagram, the area between the marginal benefit and marginal cost curves characterizes net benefits (positive ones in regions where MB>MC, and negative ones in regions where MC>MB).[2]

The relevant areas for calculating net benefits at quantities Q', Q*, and Q'' are all labelled with lower case letters in figure 2.1. The total benefits associated with quantity Q' is a+ b, which is the area under the MB curve from 0 to Q'. The total cost of Q' units is b, which is the area under the MC curve from 0 to Q'. So, the net benefits of purchasing Q' units at price P is area a—the area under the MB curve less the area under the MC from 0 to Q'. Similarly, the total benefits associated with quantity Q* is a+ b + c + d. The total cost of Q* units is b+d, so the net benefits of purchasing Q' units at price P is area a+ c. The total benefits associated with quantity Q'' is a+ b +c +d +f. The total cost of Q' units is b+ d + e + f , so the net benefits of purchasing Q' units at price P is area a + c - e.

Notice that the net benefit realized at Q* is larger than the other two, which implies that net benefits (here, consumer surplus) are maximized at Q*, because any amount less than Q* will entail missing out on net benefits analogous to area c, and any amount more than Q* will entail bearing net losses analogous to area e. A net-benefit maximizing consumer that purchases positive amounts of a good will select a quantity where his or her marginal benefits equal his or her marginal costs.

An exception to this rule occurs when the MC curve is above the MB curve for all quantities. In that case, an individual will purchase zero units of the good. At zero, there are no benefits and no costs, so net benefits are also zero. This net benefit is larger than that associated with

[2] Since the marginal benefit and costs curves are graphs of the first derivatives of their associated total benefit and total cost curves, the total functions can be recovered by taking the integral of those functions over the range of interest, except for an unknown constant. The unknown constant is assumed to be zero in these diagrams. When "fixed costs" are relevant, the unknown constant of the integral of the first derivative of the total cost function is fixed cost (e.g. as with driving to a grocery store) and the relevant total recovered by the integral without the unknown constant is total variable cost (the cost of the purchases made once there). See any calculus textbook for more on this, if interested.

any positive quantity, because all positive quantities have costs that are greater than their benefits and thus negative net benefits associated with them.[3]

Note also, that such cases (corner solutions) are not always "weird" or "unusual" cases. Rather, they are the most common case experienced in well-developed markets. For example, the typical consumer in a grocery store purchases zero of nearly all of the hundreds or thousands of the goods available, while purchasing positive quantities of only a few or few dozen things from his or her shopping list.

The continuous forms of most models can be thought of as a characterization of the "average consumer," who, of course, purchases the average amount of everything sold. Such a person does not exist, but if one is interested in the typical outcome in markets, it is a reasonable model of the average consumer's choices.

The above diagram illustrates the case where the marginal benefit curve exhibits diminishing returns, and thus it is monotonically decreasing in the quantity purchased or used. Other cases are occasionally of interest, but this type of choice setting has attracted by far the most attention by economists, and so such settings are the main focus of this section of chapter 2.

Deriving a Demand Curve from a MB Curve

Figure 2.1 implies that individuals that purchase positive quantities of goods and services will purchase the quantity where MB=MC because it is this quantity that maximizes their net benefits. This property can be used to derive Al's demand for any familiar good of interest (burgers, bread, bananas, beers, beans, beds, bras, bugles, boomerangs, bungee cords, banjo lessons, bus tickets, etc.) from his or her MB curve.

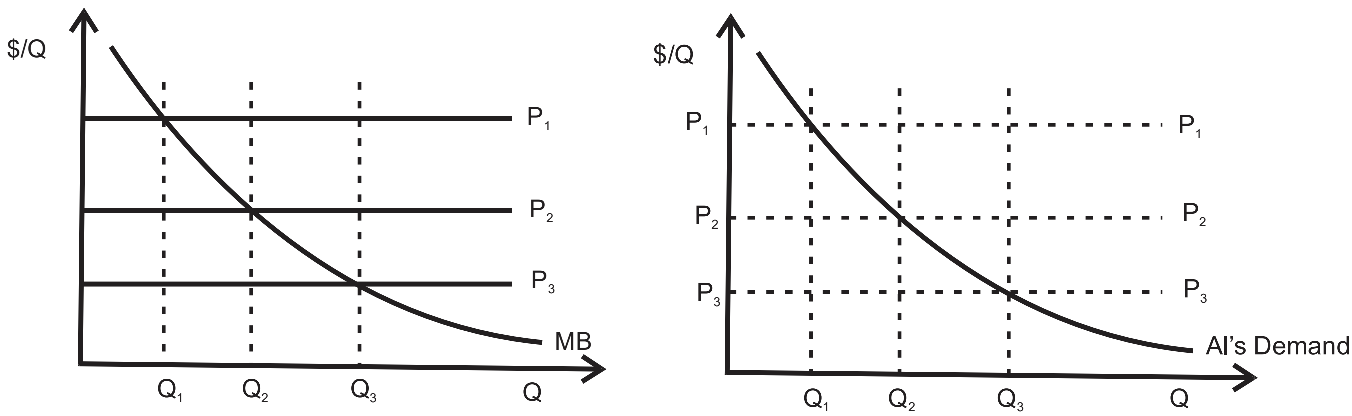
An individual's demand function is a mapping from prices into quantities purchased. Given an individual's marginal benefit curve, one can derive that individual's demand curve by (i) choosing a price, (ii) drawing in the associated marginal cost (MC) curve, (iii) finding the quantity at which the implied marginal cost curve equals the consumer's marginal benefits,

---

[3] Another exception occurs when goods are available only in discrete units. In that case the Benefit and cost functions are a sequence of points rather than curves, and individuals will purchase the last amount where the marginal benefit is larger than marginal cost. In the rare case where there is nonetheless a unit where marginal benefit equals marginal cost, the individual will be indifferent between that unit and one less unit of the good—and may simply "flip a coin" to determine which quantity to purchase.

and (iv) then plotting that price and quantity on another diagram. This process is repeated to trace out a demand curve. A complete map of Al's demand curve requires this to be done for all possible prices, but doing this for a few prices is normally sufficient to indicate the shape of an individual's demand curve. See figure 2.2a.

Figure 2.2a: A Consumer's Demand Curve



This process is illustrated for three different prices in Figure 2.2a. When a consumer is a "price taker" he or she has no bargaining power and adjusts his or her purchases to the market price. The prevailing market price is his or her marginal cost for the good of interest (ignoring travel and shopping costs—as is usually done). It characterizes the increase in cost associated with a one unit change in quantity (and also the rate at which total cost increases for infinitesimal increases in quantity).

An implication of figure 2.2a is that when prices rise, as from P_2 to P_1, the consumer will purchase fewer units of the good, $Q_1 > Q_2$. The last few units purchased under P_1 had produced positive net benefits, but under P_2 produce negative ones. This implies that individual demand curves and their associated functions are "downward sloping" whenever an individual's marginal benefit curve exhibits diminishing marginal returns.

Notice that this consumer's demand curve passes through exactly the same points as his or her MB curve in such cases. (P_1, Q_1), (P_2, Q_2), and (P_3,Q_3) are all points on both the marginal benefit curve and the individual's associated demand curve. This is true of all the other points on the demand and marginal benefit curves, whenever an individual's marginal benefit curve is monotone decreasing (is downward sloping left to right) and the MB curve is initially above the MC curve. This is the standard "textbook" case, although it is not the only possibility.

A more general implication of the geometry of figure 2.2 is that anything that affects a consumer's marginal benefits will also affect his or her demand curve and its associated demand function. Increases in marginal benefits shift the demand curves outward to the right (e.g. in the Q dimension). Decreases in marginal benefits shift their demand curve(s) inward to the left.

This inference clearly follows in every case in which the consumer's marginal benefit curve is monotone decreasing (e.g. $d^2B/dQ^2 < 0$). In that case, an individual's demand curve goes through exactly the same points as his or her marginal benefit curve.

Nonetheless, the marginal benefit function and its associated demand function are not the same functions. Marginal benefits are a mapping from quantities into benefits per unit (mapping from the horizontal axis to the vertical axis). Demand is a mapping from prices into quantities (from the vertical axis to the horizontal axis). They are inverse functions for one another.

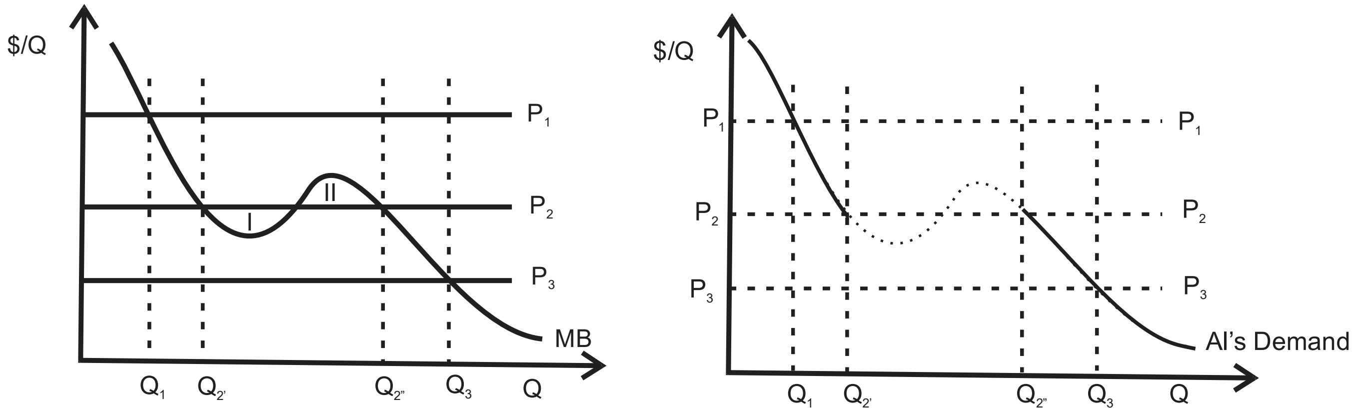
This basic model of consumer behavior can easily be extended to other walks of life and to circumstances not often modelled in microeconomic textbooks. For example, a person that is running for an elective office in a democracy may be modeled as a "net vote maximizer." Such candidates take policy positions that maximize their net votes—realizing that as they shift among policy positions, they will lose some votes (their marginal cost in terms of votes) but gain others (their marginal benefits in terms of votes).

In settings where risk or time are important, the persons may be modeled as maximizing "expected" net benefits or the "present value" of net-benefit flows, rather than realized or instantaneous benefits. Cost-Benefit analysis is grounded in extended models of net-benefit maximizing choice.

Deriving a Demand Curve from an Odd MB Curve

As mentioned above, there can be cases in which a marginal benefit curve is not monotone decreasing. Such cases often exhibit what might be called internal complementarity. As the quantity of a good increases, new possibilities occasionally emerge that cause marginal utility to increase (as with the fourth automobile tire or even numbered shoes). Figure 2.2b illustrates such a case.

Figure 2.2b: A Consumer's Demand Curve with an Odd MB Curve



Derivation of an individual's demand curve proceeds as above. However, in this case, price P_2 is chosen to make area I=II. In that case, the net benefits associated with $Q_{2'}$ and $Q_{2''}$ are exactly the same. Thus the individual is indifferent between these two quantities ($Q_{2'}$ and $Q_{2''}$) of the good of interest.

At any price higher than P_2 the individual will choose the intersection with MC on the left and at any price lower than P_2 he or she will prefer the intersection with MC on the right. Thus, there is a discontinuity in the demand curve at price P_2. (If this is not clear, calculate the net benefits associated with various quantities using the method developed in figure 2.1.)

Note that the demand curve still includes only points from the MB curve, but in this case, not every point on the MB curve is also a point on the individual's demand curve. Again, any change in the marginal benefits that the individual receives will alter their demand curve. However, in this case, only changes that affect the marginal benefits for quantities lower than $Q_{2'}$ or higher than $Q_{2''}$ are likely to do so.

Cases like this are implicitly ruled out when we assume that the net benefits functions are strictly concave. Thus, although the calculus-based derivations of demand are in a sense more general than the geometric ones, there are cases that cannot be easily analyzed using calculus-based models that are easy to characterize using geometric models.

In some cases, the neglected cases may be important, as with a normal consumer's shopping basket of goods purchased in a grocery store, and for cases in which the quantities of interest are discrete (available only in particular sizes, rather than continuous ones). Discontinuities may also partly account for the fact that some goods are available only in particular discrete sizes—although a firm's savings on packaging, spoilage, and labor (scales and package weighers) doubtless are more important determinants.

Nonetheless, students should keep those special cases in the back of their minds as we develop the standard models in the rest of this part of Part I.

The core neoclassical models are all widely taught, although they are not entirely general. They provide us with useful explanations of commonplace market phenomena, rather than every market phenomenon that may occur. They are broadly true in most ordinary cases, although they may be misleading in unusual ones. They are models of reality rather than reality itself.

The core neoclassical models provide a variety of insights that deepen our understanding of millions of individual product markets and market transactions—although they do not account for everything. That such a relatively small collection of relatively lean models can explain so much about so many markets is remarkable.

III. Concrete Functions, Net Benefit Maximization, and Consumer Demand

Geometric derivations of demand and other relationships are not as dependent on mathematically convenient assumptions as calculus-based derivations are. Calculus-based models often provide more precise results than geometry allows, can be more easily used to characterize how changes in the most relevant choice settings affect market outcomes, and their assumptions made are usually more obvious than in geometric models. They also allow more causal factors to be systematically taken into account.

Benefit functions may, for example, include income or the prices of substitutes as well as quantities purchased, which allows the effects of those variables on demand to be characterized. Such extended models explain why MB or MC curves "shift" as parameters of the choice setting other than price change. Further extensions allow quite general proofs to be developed that demonstrate that prices exist that can generate a general equilibrium for any countable number of markets, albeit with various continuity assumptions. (An example of such a proof is provided in the appendix of chapter 5.)

Calculus-based derivations also demonstrate that the mathematical methods used in the "hard" sciences and engineering can also be applied in the social sciences—although it should be admitted, often with less than entirely realistic assumptions and implications. Models based on differential equations, real analysis, and topology have also been developed by microeconomic theorists. However, most students of advanced microeconomics have taken calculus courses and understand how it can be used to characterize various properties of a function including the determination of the value of a variable (often X) that maximizes the value of a particular function (often describing how values of Y are affected by values of X). Relatively few will be as familiar with more advanced mathematical concepts and results.

The rest of this text demonstrates that calculus can be used to develop sophisticated models without spending a lot of class time introducing students to new mathematical ideas.

In the hard sciences, greater rigor is often associated with more predictive power and precision, although that is not always true in the social sciences. This is partly because the social sciences model agents that have volition, rather than inanimate objects. There are far more factors that influence individual decisions than the path of a rock. And because social outcomes are joint products of many individual decisions, even more factors are required to fully characterize them than to determine the orbit of the moon around the Earth or the Earth around the Sun.

As noted in chapter 1, Newtonian mechanics and gravity do not provide as much insight into the paths of creatures that are self-propelled and make their own independent decisions—although Newtonian factors are not entirely irrelevant. No human can leap over a tall building with a single bound or fly by moving their hands up and down in a swimming motion while on earth. On the other hand, no rock, however large or small, makes a decision about when to fall or where to land.

To illustrate how a net-benefit maximizing choice can be modelled using functions and calculus, we'll make a few assumptions about the shapes of typical total benefit and total cost functions. We'll initially assume specific forms for the functions that characterize net benefits and then move on to more abstract functions to demonstrate more general results.

Suppose, for example, that Al's total benefit function for the good of interest is B= $5Q^{.8}$. If Al can buy as much of Q as he or she wishes at price P, then Al's total cost function is just PQ, the amount paid for Q units of the good of interest. Given these assumptions, Al's net benefit function can be written as

$$N = B - C = 5Q^{.8} - PQ \qquad (2.1)$$

The first derivative of the net benefit function with respect to Q is:

$$dN/dQ = 4Q^{-.2} - P \qquad (2.2)$$

The first term ($4Q^{-.2}$) is Al's marginal benefit, and the second term (P) is Al's marginal cost.

The second derivative of the net benefit function is:

$$d^2N/dQ^2 = -.8Q^{-1.2} \qquad (2.3)$$

The second derivative of the net-benefit function is simply the slope of Al's marginal benefit curve, because the slope of the marginal cost function is zero (it does not change with quantity). It is a horizontal line with slope 0 in the Q domain $(d^2C/dQ^2 = 0)$, unless there are quantity discounts or surcharges. Since the slope of Al's marginal benefit function is less than zero for all positive quantities, a graph of Al's marginal benefit function would be downward sloping as in figure 2.1 for any $Q > 0$.

Calculus implies that **a function is strictly concave if** its second derivative is less than zero in the domain of interest. (A short overview of the geometry of concavity and few other useful mathematical concepts is provided in the appendix of this chapter.) Such functions will satisfy the geometric definition—a cord connecting any two points on the function will lie below it. Thus, equation 2.3 demonstrates that this particular total benefit function is strictly concave, which in this case also implies that the net benefit function is also strictly concave.

Strictly concave functions have at most one maximum value. The benefit function is strictly concave but it does not have a maximum value. (It is monotone increasing.) However, the net benefit function has a unique maximum value at the quantity where MC=MB, as in the figure 2.1.

The geometry behind calculus implies that a function is at a maximum or minimum whenever the first derivative of the function of interest has the value 0, and the second derivative is also negative at that point. The calculus above shows that this is true of this particular function whenever $4Q^{-.2} - P = 0$, e.g. whenever marginal benefit ($4Q^{-.2}$) equals marginal cost (P), as in figure 2.1.

A bit of high school algebra can be used to determine whether such a quantity exists.

$$4Q^{-.2} - P = 0 \qquad (2.4)$$

$$4Q^{-.2} = P \qquad (2.5)$$

$$Q = (P/4)^{-5} \qquad (2.6)$$

(Recall that raising a variable (such as Q) with an exponent (such as a) to some power (such as b), has the property that $(Q^a)^b = Q^{ab}$.)

Equation 2.6 shows that there is a quantity that maximizes net benefits for any positive price P that might exist. Indeed, equation 2.6 precisely describes what that quantity will be for every possible price. Equation 2.6 is thus Al's **demand function** for the good of interest.

If Al has the assumed net benefit function and makes decisions that maximize net benefits, then equation 2.6 would precisely characterize the quantities of the good that Al would purchase at any positive price.

The quantity that solves the "first order condition," (e. g. sets the first derivative of an "objective function" equal to zero) is often denoted with an asterisk (*), to indicate an ideal value, as with $Q^* = (P/4)^{-5}$. Note that Al's demand function is monotone decreasing in price. The slope of the demand function is $dQ^*/dP = -5(P/4)^{-6}$, which is less than 0 for all positive prices.

This example demonstrates that if we know the specific function that describes an individual's benefit function, we can derive his or her demand function.

However, in many cases, we may not know the specific benefit function, but rather believe that it belongs to a particular family of concrete functions. For example, Al's total benefit function may not necessarily be $B = 5Q^{.8}$, but is believed to belong to the family of strictly concave exponential functions, $B = aQ^b$ with $a > 0$ and $0 < b < 1$. (If "a" were not greater than 0, it would not be a good, and if b were not between zero and one, the benefit function would not be strictly concave.)

To derive Al's demand curve, we repeat the steps above given our assumption about the specific family of concrete functions Al's net benefit function belongs to. Net benefits in this case are:

$$N = B - C = aQ^b - PQ \qquad (2.7)$$

Its first derivative with respect to Q is:

$$dN/dQ = abQ^{b-1} - P \qquad (2.8)$$

The first term (abQ^{b-1}) is Al's marginal benefit, and the second term (P) is Al's marginal cost. Al is a "price taker" and thus price is a parameter of his/her choice problem rather than a control variable that Al can manipulate. The second derivative of the net benefit function with respect to Q is:

$$d^2N/dQ^2 = ab(b-1)Q^{b-2} \qquad (2.9)$$

The second derivative is negative. Recall that $b - 1 < 0$, $b > 0$, $a > 0$, and $Q > 0$; thus this entire family of net benefit functions is strictly concave.

To characterize the ideal quantity that Al should (and would) purchase, we set equation 2.8 equal to zero and solve for Q.

$$abQ^{b-1} - P = 0 \quad (2.10)$$

A couple of algebraic steps are sufficient to solve for Q*, the quantity that maximizes Al's net benefits from the product or service of interest.

$$abQ^{b-1} = P \quad (2.11)$$

$$Q^* = (P/ab)^{1/(b-1)} \quad (2.12)$$

An ideal quantity, Q^*, exists for each possible member of this family of functions (e.g. for each value of a and b). Note that Q* is always a positive quantity, and that according to equation 2.11, Al's ideal quantity always occurs where marginal benefit equals marginal cost.

Equation 2.12 characterizes the family of functions that Al's demand function belongs to for all the possible values of a and b in the assumed family of total benefit functions that are within their assumed ranges ($a > 0$ and $0 < b < 1$)).

In this case, the results point to a family of possible results rather than a unique one as in the previous example, but they all share particular properties. For example, the demand curves all slope downward. (Note that $\frac{dQ^*}{dP} = [1/(b-1)](P/ab)^{[1/(b-1)]-1}$, which is less than zero for all positive prices, because $b - 1 < 0$.)

The above results for strictly concave exponential benefit functions are useful for studies that attempt to estimate demand functions, because this family of demand functions is "linear in logs" or "log linear." If you take the log of both sides of equation 2.12, the result resembles the equation of a straight line:

$$log(Q^*) = (1/b - 1)\ log(P/ab) = (-1/b - 1)\ log(ab) + (1/b - 1)\ log(P).$$

The first term is a constant, [$(-1/b - 1)\ log(ab)$] and the second is a constant times the log of price [$(1/b - 1)\ log(P)$].

Thus, if Al's total benefit function is a member of this family of functions, then Al's demand curve for the good of interest can be estimated in logs using the conventional linear methods from econometrics given a data set for prices and purchases by Al or a group of consumers with similar benefit functions.

IV. Net Benefit Models with More General Families of Functions

The problem with results obtained from models that rely upon explicit functional forms is that they are often special cases. The function families focused on often have properties that can be easily worked out using algebra and calculus—which is to say they are mathematically very tractable functions—although they may have properties that are unique to that family of functions. (For example, not all demand functions are log linear.)

Of course, it is their "tractability" (ease of finding clear and simple solutions) that is the main attraction of the families of explicit functions used by economists. Their tractability allows one to focus attention on key features of the choice setting of interest rather than on the mathematics of the models developed. Such functions often have implications about the algebraic structure of demand and supply functions of interest that are fairly easy to estimate with well-understood econometric methods.

Models grounded in more general families of functions may also be mathematically tractable, although the results tend to be more abstract, and provide less guidance about how to estimate the model developed, although they do allow one to characterize in a more general way the relationships among choice-influencing variables.

More general characterizations of families of functions naturally yield more general results with fewer odd or idiosyncratic properties. For example, it would be surprising if all demand curves were linear in logs as implied by the exponential characterization of the total benefit function developed in the previous section—although this could well be a useful first approximation for many demand curves. Moreover, many economists are more interested in qualitative relationships—such as demand curves slope downward—than in estimates of a particular demand function, because qualitative results help explain a broad range of phenomena, rather than the behavior of a particular individual in a particular market at a particular time in his or her life.

If a general analysis implies that most demand curves slope downward, then one can use that result as the basis for a qualitative analysis of most markets, without having to estimate demand curves for each of the markets of interest. Often qualitative—intuitive—analysis is sufficient to understand, for example, roughly how a particular natural disaster or type of public policy is likely to affect a particular market or network of linked markets.

A general model of net-benefit maximizing choices by consumers in markets in which they are price takers is easy to develop. Let $B = b(Q)$ be the individual of interest's total benefit function and $C = PQ$ be his or her total cost function. Assume that function b is strictly concave (e.g. has positive first derivatives and negative second derivatives) in the domain of interest ($Q \geq 0$). Net benefits can be written as:

$$N = b(Q) - PQ \tag{2.13}$$

Differentiating with respect to Q, and setting the result equal to zero implies that net benefits are maximized when:

$$dB/dQ - P = 0 \tag{2.14}$$

Note that $d^2N/dQ^2 = d^2B/dQ^2 < 0$, since function b is assumed to be strictly concave. So, if a quantity Q exists that satisfies equation 2.14, it is the one that maximizes net benefits. The first term, as in the previous cases, is the marginal benefit of the good or service and the second is its marginal cost.

To characterize an individual's demand function using such abstract functions requires two somewhat advanced "tools" from calculus. The first is the **implicit function theorem,** which states that given a differentiable function $h(a, b, c, d) = 0$, then a function h exists such that $a = f(b, c, d)$. In other words, given a first-order condition (such as equation 2.14), it is possible to develop a function that describes how any one of the variables in the first order condition is affected by the others in the first order condition.

There are just two variables in the first order condition characterized by equation 2.14, namely, Q and P. Thus, a function f exists that can describe how Q is affected by P.

$$Q^* = f(P) \tag{2.15}$$

That function is Al's demand function for the product or service of interest—the one for which we have assumed a strictly concave benefit function.

The second tool that we'll apply is the **implicit function differentiation rule.** Suppose that we have used the relationship $h(a, b, c, d) = 0$ to characterize "a" as a function of the other variables in function h, $a = f(b, c, d)$. In that case, the derivative of function f with respect to variable b is

$$da/db = dh/db \,/\, (-\, dh/da).$$

(This rule can be derived by finding the total derivative of function h, assuming that only variables a and b change, and then applying a bit of algebra.)[4]

In most cases in this textbook, function "f" will be a first order condition. For example, if function "f" is described by equation 2.15, then equation "h" is that characterized by equation 2.14. In that case, the slope of the demand curve is:

$$dQ^*/dP = dh/dP\,/\,(-dh/dQ) = -1/-(db^2/dQ^2) < 0 \quad (2.16)$$

If an individual's total benefit function is strictly concave, then his or her demand functions are all downwards sloping. Thus, it turns out that the geometry of figure 2.2 is quite general. As long as the marginal benefit curve is downward sloping (as implied when the total benefit function is strictly concave), then demand curves derived from the net-benefit maximizing model of consumer choice are always downward sloping.

V. Utility Functions and Simultaneous Purchases of Combinations of Goods

The net-benefit maximizing model of rational consumers has clear implications and, arguably, is the basis for many intuitions about how consumer demand operates. Microeconomists often think about one market at a time. However, there are choice settings in which several goods are chosen simultaneously and individuals are constrained in what one can purchase by income, borrowing, and time constraints. Neither possibility is taken into account in the net-benefit maximizing models worked out above.

Another—and more widely used—model of rational consumer behavior assumes that individuals have a unified objective normally referred to as "utility," a concept introduced by utilitarian philosophers such as Jeremy Bentham in the late eighteenth century. Initially, utility was interpreted as a synonym for the (net) happiness or satisfaction associated with a particular pattern of behavior and outcomes.

Early utilitarians argued that all persons were utility maximizers—as do most contemporary economists. Although the meaning of utility has shifted somewhat through time, the original

[4] The implicit function differentiation rule can be derived fairly easily. Let $0 = h(a, b, c)$. The total derivative of *h* can be written as $\Delta a \frac{dh}{da} + \Delta b \frac{dh}{db} + \Delta c \frac{dh}{dc}$. Note that since h always has the value zero, its total derivative is also always zero. Its value never changes, so $\Delta a \frac{dh}{da} + \Delta b \frac{dh}{db} + \Delta c \frac{dh}{dc} = 0$. Now suppose that only a and b vary. Then $\Delta a \frac{dh}{da} + \Delta b \frac{dh}{db} = 0$ which implies that $\Delta a \frac{dh}{da} = -\Delta b \frac{dh}{db}$. This in turn implies that $\frac{\Delta b}{\Delta a} = \frac{dh}{da} / -\frac{dh}{db}$. The latter is, of course, **the implicit function differentiation rule**. If *a* were price and *b* were quantity, and *h* characterizes the first order condition of a constrained optimization problem, this equation would characterize the slope of the demand curve in a very small (infinitesimal) region around Q*, as above.

utilitarian one is still widely used.[5] When applied to consumer choice, an individual's utility is normally characterized as a monotonically increasing function of the quantities of goods and services that are at an individual's disposal, or the quantity consumed—although most goods are not literally consumed. One does not eat an automobile but may gradually use it up through wear and tear, although most folks "trade their cars in" well before they wear out (e.g. is fully consumed).

A typical general utility function for individual i is written as $U_i = u_i(A_i, B_i, C_i, \ldots)$, where $A_i, B_i, C_i, \ldots$ are the quantities of goods and services consumed by individual I, and the first partial derivatives of the utility function for each good are all greater than zero, and their second partial derivatives are negative, so that marginal utility curves for each good in the utility function exhibit diminishing marginal utility.

We'll again start with the geometry of such choices. The calculus of consumer choice follows in the next sections. Illustrations of the geometry of rational choice are limited to two-dimensions, because that is the nature of both a printed page and display screens. In order to be able to show decisions involving 2 goods, utility levels are characterized with indifference curves. Indifference curves show all the combinations of the two goods one is indifferent between. Such combinations have the same utility level. In effect, indifference curves provide a topographic map of a consumer's utility function. Hikers are familiar with topographic maps if they've planned to climb mountains and wanted to know how high and steep the way to the top of a mountain is.

Indifference curves similarly allow a three-dimensional "utility mountain" to be reduced to a two-dimensional diagram. Although conceptually indifference functions can take any shape

[5] Most utilitarians were and continue to be more interested in appraising the merits of alternative public policies and institutions than in describing consumer behavior. From the utilitarian perspective, one policy is better than another if it increases the sum of utility (happiness or life satisfaction) in a community or the average level of utility in a community more than another. Insofar as these included economic policies, various ideas about market networks were developed by utilitarians and several utilitarians wrote economic textbooks including James Mill (1821), John Stuart Mill (1848), and William Stanley Jevons (1871), who developed the idea of a utility function, as noted above. Many of the most remembered "economists" from the late nineteenth century were utilitarians. The idea of an indifference curve was worked out by Edgeworth (1881) although not drawn until Pareto (1906). Their ideas about aggregate utility still characterize most of the field of "welfare economics" today. (Welfare economics is taken up in chapter 19.)

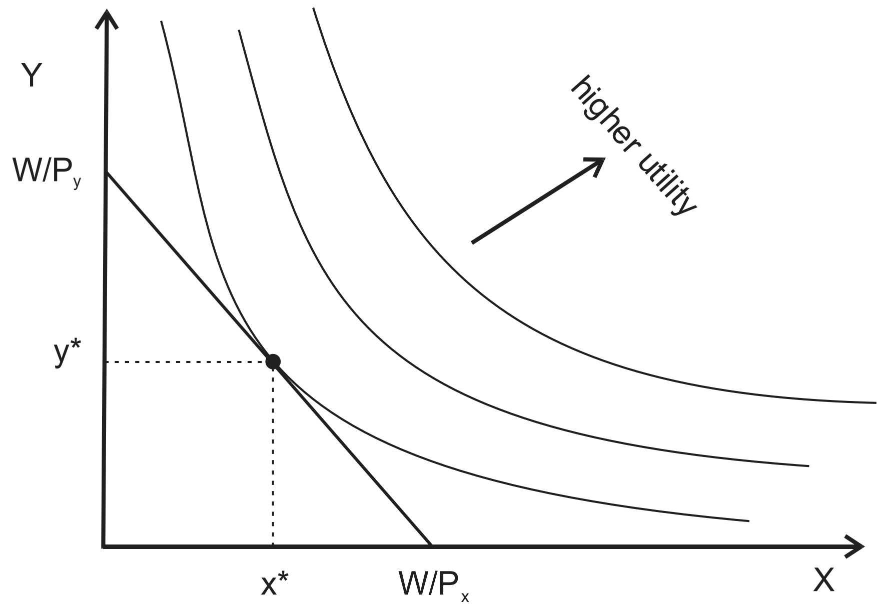
that a mountain can, the usual assumption about utility functions limits them to a series of more or less C-shaped curves that represent higher utility levels as one moves to the northwest (upper right) of the diagram. More of each or both goods is normally assumed to be better, e.g. to generate higher utility. Thus indifference curves to the upper-right of most indifference curve diagrams are "higher" (denote higher levels of utility) than curves to the lower left. C-shaped curves exhibit diminishing marginal rates of substitution between the various "bundles" or combinations of the two goods being chosen.

However, consumers cannot choose any combination included on such a diagram, because they are constrained to pick ones that he or she can afford. Only combinations that lie within their opportunity or budget set are feasible. When consumers are price takers and every unit of the goods of interest have constant prices, their budget line is a downward sloping line that characterizes the upper bound of a triangular opportunity set as shown below in figure 2.3.

Figure 2.3 depicts a typical consumer's choice (Al's) between goods X and Y for the case in which the price of good X is P_X and the price of good Y is P_Y and the consumer has amount W to spend on the two goods in the period of interest. The assumptions about prices and the total amount to be spent on the two goods implies that Al has a budget constraint that can be represented algebraically as $W = P_X X + P_Y Y$, which appears in figure 2.3 as a diagonal line running from W/P_Y on the vertical axis to W/P_X on the horizontal axis. That line is called a budget line or a budget constraint.

The end points characterize the maximum quantity of good X or Y that can be purchased if Al spends all of his or her money on just one of the goods. The slope of the budget line is determined by the relative prices of the two goods, which in this case is $-P_X/P_Y$.

Figure 2.3: The Geometry of a Utility Maximizing Choice



The highest indifference curve that can be reached is one that is tangent to the budget line. At that combination of goods X and Y, Al has achieved the highest utility that is feasible given the amount that he or she has to spend on the two goods, W.

A demand curve remains, as above, a mapping of prices into quantities purchased, but the derivation is a bit different than in the net benefit maximizing case. To derive a demand curve, one of the prices is varied and the other is held constant as is the consumer's budget (W). Changes in the price of interest change the budget set and thereby affect the consumer's choice of both goods.

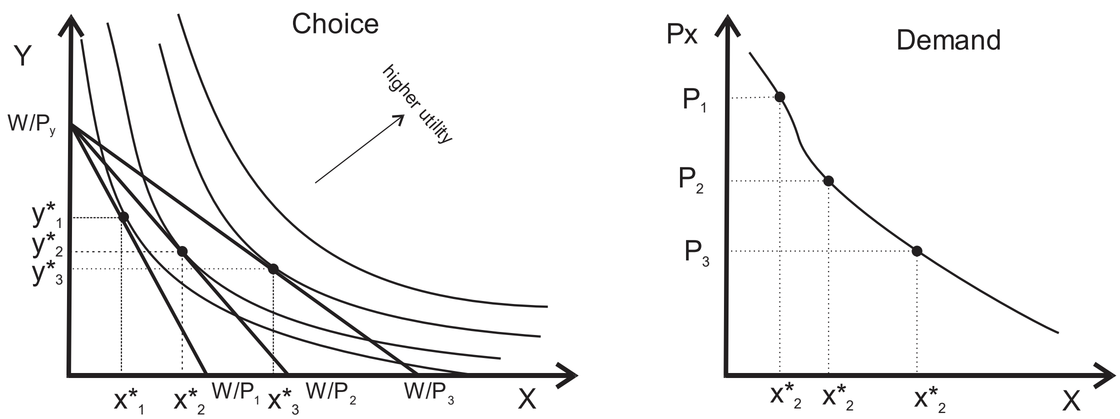
To illustrate the process, the demand for good X is derived below. To do so, various prices for X are tried and the price and quantity of good X purchased at the price are plotted in the diagram to the right to trace out Al's demand curve for good X. In each case, Al purchases the quantities of X and Y that maximize his or her utility, given the "new" price for good X.

Figure 2.4 illustrates how this process operates geometrically. Three prices of good X are selected with P_1 the highest and P_3 the lowest. This is evident in that the highest quantity that Al could purchase of good X, (W/P_X), increases as prices fall from P_1 to P_2 to P_3.

Every time the price changes, Al changes his "bundle" of X and Y, because his budget line and opportunity set change. Note that both X and Y change as the price of X changes, not simply X as might be suggested by a straight-forward net benefit maximizing model.

The resulting demand curve is "qualitative" because no specific values for W, P_Y and P_X have been specified, nor has a scale for the two axes. If those had been specified, particular values for X and Y would have emerged from a careful examination of diagram, given the shapes and positions of the indifference curves.

Figure 2.4: Indifference Curves and Demand



In the case depicted, demand is downward sloping (monotone decreasing) although this is not always the case with indifference-curve based analysis. It is, however, by far the most common type of demand curve found in empirical research.

However, it bears noting that a demand curve derived from indifference curves is not necessarily linear or smoothly downward sloping as often assumed in empirical work. Note also that the geometry of this derivation of the demand curve implies that an individual's demand curve (or demand function) for X shifts if wealth (W) or the price of the other good (P_2) changes.

VI. Exponential Multiplicative Utility Functions and Consumer Demand

Mathematical models of utility maximizing choices can take account of more than two goods, although, perhaps surprisingly, two-good models are sufficient to illustrate the most common results associated with deriving demand curves using calculus. There are two calculus-based methods that can be employed to derive a demand curve from a utility curves. The first is called the Lagrange method (named after the French mathematician that worked out this method for constrained optimization) and the second can be called the substitution method. (Other methods that you might see in other textbooks are generally variations on these two.)

As in the section on net-benefit maximizing choice, we'll initially assume that a specific concrete functional form for a utility function and then assume a more abstract utility function with a specific shape. In both cases, the utility function will be assumed to be strictly concave and represent choices over two goods. Being "goods," more of either or each increases utility. Thus, the ideal combination (utility maximizing combination) for any individual will lie along his or her budget constraint. We will again refer to the individual of interest as Al (short for Allen or Alice, so Al can be used as a name for either a male or female person.)

The "work horse" of this section of chapter 2 is a generalization of the Cobb-Douglas (1928) family of functions. Cobb and Douglas showed that a function of the form $U = X_1^a X_2^{1-a}$, with $0 < a < 1$, and X_1 and X_2 being two goods. This functional form has a variety of useful mathematical properties that are convenient for constrained optimization, some of which are a little odd. Nonetheless, it turned out to be useful as a base for estimating many economic relationships.

In this book, we'll more often use a generalization of that relationship that drops the assumption that the sum of the exponents is 1, $U = X_1^a X_2^b$ with $0 < a < 1$ and $0 < b < 1$. That family of functions can be termed "multiplicative-exponential" functions, a descriptive but somewhat awkward rubric.

Either type function can be used for as many goods as one might be interested in, but normally we'll focus on two good choices. This is partly because this is sufficient to generate most of the interesting implications of consumer choices and partly because it simplifies the mathematics a bit.

A Short Introduction to Partial Derivatives and the Lagrange Method of Constrained Optimization

To derive demand curves using even relatively straightforward generalizations of the exponential model of net-benefit based demand developed above, requires somewhat more sophisticated tools from calculus to be employed, namely partial derivatives and the Lagrangian method. For those whose calculus classes mostly focused on functions with a single control variable, such as X, a short introduction to the idea of a partial derivative is useful. That is followed by a short introduction to the Lagrangian method.

A **partial derivative** characterizes the effect of a change in one variable of a multi-variate function on the function's value, holding all the other variables constant. For example, suppose that Al has a three-good multiplicative utility function. The three goods may be X, Y, and Z, and Al's utility function may be $U = X^a Y^b Z^c$. In that case, Al has three control variables, X, Y, and Z—the goods that he or she may purchase. Note that the exponents and initial term are variables in the mathematical sense, but they are not control variables, because they are assumed to be unalterable aspects of Al's utility function for the purposes of the model. The partial derivatives of the utility function with respect to the control variables are usually the main focus of the analysis neoclassical models, because tastes (utility functions) are normally assumed to be constant during the period of analysis.

Partial derivatives focus on the effect of small changes in one of several variables that determine the function's value. Each variable may have an effect on a function's value and their individual effects (partial derivatives) are computed one at a time.

The partial derivative of Al's utility function with respect to X is $\frac{dU}{dX} = aX^{a-1}Y^b Z^c$. In effect, the Y and Z terms ($Y^b Z^c$) are treated as constants for the purposes of finding the partial derivative with respect to X. The partial derivative of U with respect to X in this case is the effect of a small change in X on Al's utility level, other things being equal. The other two partial derivatives of Al's utility function with respect to Al's control variables (X, Y, and Z) can be found in the same way by holding the other variables than the control variable of interest constant:

$$\frac{dU}{dY} = bX^a Y^{b-1} Z^c \text{ and } \frac{dU}{dZ} = cX^a Y^b Z^{c-1}.$$

The **Lagrange method** for characterizing a constrained optimum was developed in 1764 by Joseph Louis Lagrange, one of many innovative mathematicians that can be regarded as a genius in that field of research. His method is relatively straight forward and widely used in introductory mathematical economics courses.

Given a strictly concave differentiable objective function, such as U= $\mathrm{u}(\mathrm{X}, \mathrm{Y}, \mathrm{Z})$, and a constraint in a form equal to zero, such as a three good budget line, $0 = \mathrm{W} - \mathrm{P}_X\mathrm{X} - \mathrm{P}_Y\mathrm{Y} - \mathrm{P}_Z\mathrm{Z}$, the combination of X, Y, and Z that maximizes U can be characterized as follows.

- First, form the "Lagrangian function", which combines the objective function and the constraint function multiplied by parameter λ, as with:

$$\mathcal{L} = \mathrm{u}(\mathrm{X}, \mathrm{Y}, \mathrm{Z}) - \lambda(\mathrm{W} - \mathrm{P}_X\mathrm{X} - \mathrm{P}_Y\mathrm{Y} - \mathrm{P}_Z\mathrm{Z}) \quad .$$

- Second, take the partial derivatives of the Lagrangian function with respect to each of the control variables (here X, Y, and Z) and also the partial derivative with respect to the Lagrangian multiplier, λ.
- Third, set all of the partial derivatives equal to zero. The resulting system of equations describes the optimal values of the "control variables" (here, X, Y, and Z) given the constraint that must be satisfied (here the budget constraint).

In some cases, one will be able to solve for explicit functions that characterize the ideal levels of each of the control variables given the constraint. In others, one cannot, although some light is cast on the properties of the associated ideal values of the control variables by the individual first order conditions (partial derivatives at the ideal values). It turns out that the family of multiplicative exponential functions is a case in which solutions can be computed using a bit of algebra on the system of partial derivatives that emerge from the Lagrange method.

Deriving Demand Curves from Multiplicative Exponential Utility Functions Using Lagrange's Method

With the Lagrange method in mind, we are now in position to characterize an individual's demand curve from a utility function and budget constraint using calculus. A two-good utility function is sufficient to demonstrate the technique and some of the main results, although utility functions can include any number of goods and services.

Let Al's utility function be $U = aX^bY^c$ and his or her budget constraint be $[\mathrm{W} = \mathrm{P}_X\mathrm{X} + \mathrm{P}_Y\mathrm{Y}]$, where W is the amount that can be spent on X and Y in the period of interest and P_X and P_Y are their prices. Given these assumptions, the associated Lagrangian function is:

$$\mathcal{L} = aX^bY^c - \lambda[\mathrm{W} - \mathrm{P}_X\mathrm{X} - \mathrm{P}_Y\mathrm{Y}] \qquad (2.16)$$

Differentiating with respect to X, Y, and λ, and setting the results equal to zero generates the following system of first order conditions that characterizes Al's ideal levels of X and Y, given his or her budget constraint.

$$\frac{\mathrm{dL}}{\mathrm{dX}} = \mathrm{abX}^{b-1}\mathrm{Y}^c - \lambda\mathrm{P}_X = 0 \qquad (2.17a)$$

$$\frac{\mathrm{dL}}{\mathrm{dy}} = \mathrm{acX}^{b}\mathrm{Y}^{c-1} - \lambda\mathrm{P}_Y = 0 \qquad (2.17b)$$

$$\frac{dL}{d\lambda} = \mathrm{W} - \mathrm{P}_X\mathrm{X} - \mathrm{P}_Y\mathrm{Y} = 0 \qquad (2.17c)$$

Note that the first term in equations 2.17a and 2.17b are the marginal utility functions for goods X and Y respectively (the marginal benefit from consuming X and Y, respectively). The marginal utility of X, for example, is $\text{abX}^{b-1}\text{Y}^{c}$. It is partly determined by its own consumption level (X) and partly by the consumption level of Y. The more of good Y that is purchased and consumed, the greater is the marginal utility of X. The same is true for good Y. When the exponents are positive, the products purchased are "goods," (more is better) and in this and many other conventional models, the goods are also what might be referred to as gross complements. Consumption of each of the other good(s) increases the marginal utility from the good of interest. This is true of the entire family of multiplicative exponential utility functions. The last term in both equations is lambda times the marginal cost of those goods, which in this case is their price in money terms. The last first order condition is the budget constraint. All three equations are implications of the initial model of Al's choice setting and not themselves assumptions.

A bit of algebra allows the demand curves for both X and Y to be derived from the Lagrangian first order conditions, and some interesting properties to be characterized.

The following steps are often used when working with this family of functions and constraints. They may be regarded as a series of algebraic "tricks" that allow several important properties of the system of first order conditions (partial derivatives) to be characterized—including relatively simple solutions for the demand functions for X and Y.

First, add λP_X to both sides of eq 2.17a and add λP_Y to both sides of equation 2.17b. Second, divide the lefthand side of first resulting equation by the lefthand term of the second, and the righthand term of the first equation by the righthand term of the second to obtain:

$$\frac{abX^{b-1}Y^{c}}{ac\ \ ^{b}Y^{c-1}} = \frac{\lambda\text{Px}}{\lambda\text{Py}} \tag{2.18a}$$

Note that the numerator and denominator on the left include many of the same terms and can be simplified. A bit of basic algebra allows equation 2.18a to be reduced to:

$$\frac{by}{cX} = \frac{\text{Px}}{\text{Py}} \tag{2.18b}$$

This is the tangency condition between the indifference curves in the XY plane and the budget constraint, as illustrated in figure 2.3. The term on the left is the slope of an indifference curve. The second term is the slope of the budget line. The slopes of an indifference curve and budget line are the same at a point of tangency.

Next, in order to find the demand function for X, we need to isolate X by solving equation 2.18b for X.

$$X = \frac{bPy}{cPx}Y \tag{2.18c}$$

Next, we need to take account of the third first order condition, e.g. the budget constraint. To do so, we use the constraint (equation 2.17c) to characterize Y in terms of X, which in this case is: $Y = [W - P_X X]/P_Y$. Substitute that relationship into equation 2.18c for Y and simplify:

$$X = \frac{bPy[W - PxX]}{cPxPy} = \frac{bW}{cPx} - \frac{bPxX}{cPx} \tag{2.18d}$$

Next, we add the positive of term with X in it on the right (the last term) to both sides.

$$X + \frac{bPxX}{cPx} = \frac{bPyW}{cPx} \tag{2.18e}$$

Isolate X on the lefthand side of the equation, and simplify:

$$X\left[1 + \frac{bPx}{cPx}\right] = \frac{bPyW}{cPx} \quad \xrightarrow{yields} \quad X\left[\frac{cPx + bPx}{cPx}\right] = X[\frac{c+b}{c}]$$

Thus, $X[\frac{c+b}{c}] = \frac{bW}{cPx}$ (2.18f)

Cross-multiply to isolate X on the lefthand side, then simplifying yields:

$$X = \frac{bW}{cPx}\left[\frac{c}{c+b}\right] \quad \xrightarrow{which\ implies} \quad X^* = \frac{b}{(c+b)}\frac{W}{P_X} \tag{2.19}$$

A similar series of steps yields:

$$Y^* = \frac{c}{(c+b)}\frac{W}{P_Y} \tag{2.20}$$

Note that demand curves for X and Y are remarkably simple for this family of functions, which at least partly accounts for this family of functions' widespread use to illustrate economic principles using calculus and the Lagrange method.[6]

[6] In the Cobb-Douglas special case, $b + c = 1$, and the demand curves are slightly simpler, $X^* = bW/P_X$ and $Y^* = (1 - b)W/P_Y$.

Al spends fraction $\frac{b}{(c+b)}$ of his or her wealth on good X and fraction $\frac{c}{(c+b)}$ on good Y. To find the specific quantity of X or Y purchased, we would need to know the price of the good of interest, the consumer's wealth and the exponents of his or her utility function. The price of the other goods is irrelevant for this specific family of utility functions.

Note that these demand curves are both downward sloping. The first derivative of Al's demand for X is $-\frac{bW}{P_X^2} < 0$. Note also that each implies that Al spends amounts on the two goods that are proportional to the exponents of the utility function.

One of the odd and surprising properties of this family of utility functions is that the demands for each good are independent of the prices of the other goods. Another is that when income (W) increases, expenditures on both goods increase proportionately. Thus, the income expansion paths for persons with utility functions from the multiplicative exponential family of functions are all straight lines, with a slope that varies with the relative sizes of the exponents. All goods are normal goods within this family of utility functions.

All these are rather "special" properties, but they allow one to use fairly straightforward techniques from econometrics to estimate the relevant demand and Engel's functions. If utility functions are from this family of utility functions, one can estimate demands for single goods without paying much attention to the prices of all other goods that might be purchased. And, perhaps surprisingly, it turns out that for many goods, the implications of this family of utility functions provide reasonably good approximations of both market demands for many goods and services, and also for the relationship between an average consumer's income and his or her allocation of income of those goods.

However, this is not true for all goods or other types of utility functions. There are cases in which the prices of other goods clearly affect both demand and expenditures. Substitutes

---

Students should now repeat these derivations on their own—without using the text, except to help overcome roadblocks. Once, the simplicity of the solution is known, a few students with a bit of algebraic skill and intuition will be able to derive this result through similar steps. However, most students will find that the derivation outlined above provides the most straightforward series of steps that will generate the surprisingly simple algebraic expressions for the demand functions associated with the family of multiplicative exponential utility functions.

and complements exist for most goods. And there are goods for which the relationship between income and expenditures is non-linear. Inferior and superior goods also exist.

VII. Characterizing Individual Demand Functions from More General Utility Functions

On the Usefulness of Strictly Concave Utility Functions

As in the case of net benefit maximizing models of the demand for goods and services (and most other things), utility-function-based analysis can draw from a very broad range of concrete functions and function families. There are, however, mathematical constraints on the types of functions that can potentially characterize utility if one wants to use calculus to characterize individual choices. First, the utility functions must be continuous. Second, they must be differentiable. And, third, they must have the property that a unique "best choice" exists if one wants to characterize a demand function. The latter generally requires strict concavity.

A demand function is a one-to-one relationship between prices and amounts purchased. Given an individual's utility function, every price induces a unique quantity to be purchased by the consumer of interest. Thus, for the purpose of characterizing that relationship, it is important that there be a unique utility-maximizing choice of bundles of goods associated with every combination of prices. Strictly concave utility functions have this property when the opportunity set is a convex set. **A convex set** is a set in which any line segment connecting two points within the set is entirely contained in the set of interest, which is true of both the budget line and opportunity set bounded by a budget line when consumers are price takers (e.g., individuals who adapt to prices over which they have no control). This assumption normally includes the restriction that there are no quantity discounts in order to assure that the budget set is convex.[7]

In addition to strict concavity, it is normally assumed that utility functions are at least twice differentiable so that second derivatives and cross partial derivatives can be characterized.

---

[7] **A function** is a mapping from a set into a unique point in another set, often in the set of real numbers. Thus, a mapping from one set into another set (composed of several points) would not be a function. In cases in which multiple bundles yield the same maximal utility levels (given the constraints), the result would not be a demand function but a demand correspondence. Price theory can be worked out for correspondences as well as functions, as was shown by Debreu (1959) among many other works in general equilibrium theory. However, one cannot use ordinary calculus for doing so.

Although all these assumptions somewhat reduce the generality of the models constructed, they usually do so at a modest cost.[8]

Using an Abstract Strictly Concave 2-Good Utility Function to Characterize an Individual Demand Function

A general abstract utility function has the form U = u(A, B, C, … X, Y) where the arguments of the function (A, B, C…) are all quantities of a good or service. Because they are goods, partial derivatives with respect to the quantity of each good is greater than zero (otherwise the arguments would not be goods), and their second derivatives are negative (which characterizes diminishing marginal utility and also contributes to the strict concavity of the utility functions). These properties have been routinely assumed since the marginal "revolution" in economic theory of the late nineteenth century. Their prices can be thought of as a ($P_a, P_b, P_c, \dots P_x, P_y$). A typical consumer's budget constraint is thus: $W = P_aA + P_bB + P_cC \dots + P_xX, +P_yY)$, where W is the wealth or amount to be spent on the goods of interest (those in the budget constraint) in the period of interest..

One can characterize a consumer's choice using the Lagrange method of the previous section. The result is a system of equations with a partial derivative (first order condition) for each of the control variables (A, B, C, ….), plus one for lambda (the Lagrangian multiplier). However, no neat solutions will be possible, and determining the effects of prices and income on the quantities demanded require the use of matrix methods.

Matrix methods and results of such analyses tend to be "messy" and cumbersome, and the associated mathematics shed little new light on the core properties of demand relationships beyond those associated with two good versions of general models. Consequently, they are rarely included in theoretical microeconomics research papers. They do, however, show that it is possible to model such broad choice settings using calculus-based mathematical methods.[9]

---

[8] It is interesting to note that none of these assumptions are necessary for geometric models—although such models cannot be depicted for more than two goods. Thus, for "non-standard" purposes, geometric models are often easier to construct and analyze than calculus-based models.

[9] For an overview of the required matrix methods, consult a mathematical economics textbook on bordered Hessian matrices and Taylor's matrix rule.

The Substitution Method

A general abstract utility function for two-good choice settings has the form $U = u(X, Y)$ with $\frac{dU}{dX} > 0$ with $d^2U/dX^2 < 0$ and $\frac{dU}{dY} > 0$ with $d^2U/dY^2 < 0$. The associated budget line is the same as in the previous section, $W = P_X X + P_Y Y$. The best combination of X and Y to purchase can again be characterized with the Lagrange method, which again will require the use of matrix methods to undertake comparative statics (analysis of how changes in prices or income affect demand).

However, in the two good case, there is another method that is mathematically tractable and often less cumbersome than the matrix methods that I call the **substitution method**. If one is interested in a consumer's demand for X, first solve the budget line for Y as a function of X. The budget line implies that $Y = (W - P_X X)/P_Y$ along the upper bound of the utility function. Since X and Y are both goods, it is along this line that the optimal combinations will be found. Substitute this variation of the budget constraint for Y into the utility function to create a composite utility function that evaluates utility levels along the budget constraint.

$$U = u(X, (W - P_X X)/P_Y) \qquad (2.21)$$

Notice that written in this way, the utility function has only one control variable, X. In the standard two-good model, once one has chosen X, Y is also determined. The budget constraint implies that to get more of X requires sacrificing more of Y and vice versa. In a multi-good model, the same tradeoff occurs, but among a broader range of goods. The constrained utility-maximizing model implies that opportunity costs are associated with every decision in which there are binding constraints. The nature of that opportunity cost is developed next..

To characterize the ideal level of X for given prices and income (or wealth), differentiate 2.21 with respect to X and set the result equal to zero. Because the quantity of good Y is now a characterized as a function of X, as X changes so will Y. Thus, the composite differentiation rule has to be applied the second argument in the utility function.[10]

$$\frac{dU}{dX} = \frac{\delta U}{\delta X} + \left(\frac{\delta U}{\delta Y}\right)(-P_X/P_Y) = 0 \equiv H \text{ at } X^* \qquad (2.22)$$

---

[10] . The composite differentiation function rule states if Y=g(z) and z=f(x) or Y=g(f(x)), $\frac{dy}{dx} = \frac{dg}{dz}\frac{dz}{dx}$, which is sometimes abbreviated as Y'=g'z'. Similarly, if Y=g(X, f(X)), it implies that $\frac{dy}{dx} = \frac{dg}{dX} + \frac{dg}{dz}\frac{dz}{dx}$. In the latter case, each argument containing X has its own partial derivative, and one simply adds them together to get the overall partial derivative. We'll be using this method of differentiation repeatedly in this text, so spend some time mastering it.

Notice that the first term ($\frac{\delta U}{\delta X}$) is the subjective marginal benefit of good X and the second term, $\left(\frac{\delta U}{\delta Y}\right)$ (P_X/P_Y), is the subjective marginal cost (the opportunity cost) of X from the reduction in the consumption of good Y associated with any increase in X—holding prices and income constant. Note also that both marginal cost and marginal benefit are in terms of utils here rather than dollars or some other currency as in the net benefit maximizing model. The marginal utility parts of those terms imply that the marginal costs and benefits may vary among individuals because their tastes are different (e.g. their utility functions are different).

Both the marginal benefits and marginal costs are in terms of "utils" rather than dollars in this model, but the relationship implied is otherwise similar to the relationship characterized by Figure 2.1. In this case, however, the marginal cost curve is upward sloping rather than horizontal because of diminishing marginal utility. The opportunity cost of consuming more X is consuming less Y. Its "marginal cost" is the lost utility from reductions in good Y. As additional X is consumed, one sacrifices more and more valuable units of Y because of the diminishing marginal utility of Y. Thus, marginal opportunity cost increases as X increases. (This property is not intuitively obvious until you think about it a bit.)

Although this first order condition is abstract, it is surprisingly easy to use it to characterize Al's demand functions for goods X and Y, and the effect of changes in the price of Y or income on demand.

The **implicit function theorem** states that given a function h(A, B, C) = 0, there exists another function that characterizes the relationship of each variable in terms of the others, as with A = f(B, C). Note that the first order condition (equation 2.22) is an instance of function h at X*. It has the value zero. Thus, there exists a function f that characterizes the ideal purchases of X as a function of the other parameters of Al's purchasing decision.

$$X^* = f(P_x, P_Y,\ W) \tag{2.23}$$

That function is, of course, Al's demand function for X. Al's demand for X varies with the prevailing market prices of X and Y, and his or her wealth (amount that Al has to spend on the two goods in the period of interest).

Al's demand function for Y is similar. It can be written as $Y^* = [W - P_x f(P_x, P_Y,\ W)]/P_Y$, or simply as $Y^* = g(P_x, P_Y,\ W)$. Both are ways to characterize Al's purchases of good Y as a function of the price of Y and other parameters of the choice setting, here W and P_x. Al's demand function for both goods varies with the price of the good of interest and the price(s) of the other good(s) and his or her wealth.

Using the Implicit Function Differentiation Rule to Characterize the Slope of an Abstract Demand Function

The **implicit function differentiation rule** provides a method of characterizing the derivative of equations such 2.23 that are based on the implicit function theorem. It states that the new implicit function (here, f) has derivatives that can be characterized with derivatives of the original function (equation 2.22) as long as the original function is differentiable. The utility function was assumed to be twice differentiable (e.g. to have second derivatives), so, the first order condition characterized by equation 2.22 is differentiable (by assumption). In that case, implicit function differentiation rule implies that the derivative of $X^*$with respect to P_x, P_Y, or W can be characterized. For example, it implies that the slope of the demand curve is $\frac{dX^*}{dP_x} = \frac{dH}{dP_x} / -\frac{dH}{dQ}$, where H is equation 2.22 in this case, but in general is the "zero function" upon which the implicit function is based.

To characterize the slope of Al's demand function in more detail, it is important to remember that each of the marginal utility functions in equation 2.22 has the same arguments (X and (W-PxX)/Py as the original (parent) function. Each of these variables includes X. Thus, derivatives of the partial derivatives with respect to X all include two terms for each partial derivative. This makes the derivatives more complex than one might guess from equation 2.23.

The implicit function differentiation rule implies that the slope of Al's demand for X is:

$$\frac{dX^*}{dPx} = \frac{\frac{dH}{dPx}}{-\frac{dH}{dX}} = \frac{\frac{dU^2}{dXdY}\left(\frac{-X}{Py}\right)+\frac{dU^2}{dY^2}\left(\frac{X}{Py}\right)\frac{Px}{Py}-\frac{dU}{dY}\frac{1}{Py}}{-[\frac{dU^2}{dX^2}+\frac{dU^2}{dXdY}\left(-\frac{Px}{Py}\right)+\frac{dU^2}{dYdX}\left(-\frac{Px}{Py}\right)+\frac{dU^2}{dY^2}(\frac{Px}{Py})^2]} \qquad (2.24)$$

This equation is rather abstract and complex. However, it turns out that most of the component terms have signs that are predetermined by the assumption that both X and Y are goods and subject to diminishing returns.

This implies that the first derivatives with respect to X and Y are both greater than zero and that their respective second derivatives are both less than zero. If we also assume that the cross partial is greater than or equal zero—e.g., $\frac{dU^2}{dXdY} \geq 0$ –-then this complex expression has a sign that can be unambiguously determined. It is less than zero!

To see this, first, notice that the term inside the brackets in the denominator is simply the second derivative of the utility function with respect to X, which has to be negative if the utility function is strictly concave. Both second derivatives are negative (there is diminishing marginal utility, by assumption). If $(\frac{dU^2}{dXdY} > 0$, the two terms that include cross partials are also negative because they are multiplied by a negative term $\left(-\frac{Px}{Py}\right)$. So, the assumptions of diminishing marginal utility and the assumption that cross partials are positive (or zero) are sufficient to assure strict concavity of the utility function—given a linear budget constraint.

The negative sign in front of the bracket, assures that the denominator has a positive sign. Thus, the sign of the slope of A's demand curve is determined by the numerator.

Notice that each term in the numerator has a value less than zero. The first term is a positive number (the cross partial) times a negative number, and so overall is negative. The second term is also negative (the second derivative of U with respect to Y times a positive number). The third term is a positive (marginal utility is greater than zero for all goods) times a positive (the ratio of two prices), and has a negative sign in front of it and so is also negative overall.

Thus, Al's demand function for X is downward sloping—the quantity of X demanded increases as its price falls. The assumption of concavity and either zero or positive cross partials is sufficient to assure that demand curves slope downward.

Negative Cross Partials?

If cross partials are negative, (if the two goods are gross substitutes) it is possible that the opposite result may occur, even if the utility function is strictly concave. Intermediate microeconomics courses often note, geometrically, that if the "income effect" is negative, it can be larger than the "substitution effect." Such effects require relatively large negative cross partials to occur, in which case there may be upward sloping sections of a person's demand function. However, limited income and wealth imply this cannot be the case throughout a demand function—and few if any such cases have been observed in empirical research on demand functions.

Many concrete utility functions are separable, which occurs when the cross partials have the value of zero. Using a separable utility function simplifies the associated calculus and algebra, as with $U = X^a + Y^b$, with a and b less than 1 and greater than zero.

For most purposes, utility functions with relatively large negative cross partials can be ignored, because the evidence implies that most demand curves are downward sloping in prices—with the exception of cases in which price is used by a consumer as an index of quality.

Market Demand Curves

Market demand functions are taken up in Chapter 4, the chapter on the neoclassical theory of price determination. However, a brief preview is useful at this point. Given individual demand functions, market demand functions are simply the sum of individual demands for the product or service of interest. For example, if there are N identical individuals purchasing good X, the market demand is simply $Q_X^d = Nf(P_x, P_Y,\ W)$. In that case, it is obvious that the variables that affect market demand, apart from the number of identical consumers, are exactly the same as those that affect the individual consumers.

If each consumer's demand is downward sloping, then the sum will also be downward sloping.[11] Market demand will be affected by the prices of other goods and consumer wealth, and so forth. When consumers differ, this remains true as shown in chapter 4, but is just a bit more complicated.

This characterization of market demand provides an example of how methodological individualism can explain social phenomena. A market demand function is a social phenomenon, but it is ultimately determined by individual demand curves.

The theory of demand is one of the most important advances provided by neoclassical economics. Classical economics had stressed the linkages between prices and production costs in the long run, and largely ignored demand. Those relationships were not discarded by neoclassical economics but turn up in the neoclassical model of supply, which is developed in the next chapter and in Chapter 5.

---

[11] This is straightforward to demonstrate. Suppose that good X is the good being analyzed and that each consumer of X has a strictly concave utility function with positive cross partials. In that case each person's demand curve will slope downward as developed in this section. Let Q_i be the demand function of the i'th person. Suppose there are N different consumers in the market. The overall market demand is the sum of the individual demand curves, $Q = \sum_{i=1}^{N} Q_i$. The slope of the market demand function is dQ/dP $=\sum_{i=1}^{N} \frac{dQ_i}{dP} < 0$. Note that every term in the sum is less than zero, as shown in this section, and thus, the overall sum is also less than zero. If each consumer's demand function is downward sloping in price, then the market demand for X also necessarily slopes downward.

VIII. Some General Conclusions

The development of an internally consistent and clear theory of consumer demand for the products sold in markets is one of the most important developments in microeconomics. It plays a role in virtually every model of markets ever developed. In addition, the same logic can be used to model a variety of non-market behaviors. It provides a very general model of human behavior and also for the behavior of groups of unaffiliated individuals.

The chapter began with two notions of utility. The first and the original one, was utilitarian, with roots that go back at least as far as Jeremy Bentham writing in the late eighteenth century. That theory posited that people actually do attempt to maximize utility—it is part of human nature. The second one, was developed by Paul Samuelson, who noted that as long as choices are internally consistent (for whatever reason), it is possible to characterize their behavior with a utility function.

In both cases, in domains where behavior was rational in either sense, utility functions could be used to characterize both human volition and how choices and actions would change if their choice settings change. Consumer decisions are one instance of this. If the price of a good increases, a broad class of utility functions implies that consumers will purchase less of a good—i.e., that incentives matter and demand functions generally slope downward.

Other very similar models of human decision making can be developed for household production, for investment and savings decisions, for the major and course-selection choices of students, for political choices, and for decisions about war and peace, and even with respect to the development and use of ethical and normative theories in one's own life as shown in Parts II and III of this text, and in a good deal of applied economic research published in economic journals.

The theory of demand has proven to be among the most generalizable of the core models and results of neoclassical economics.

The "rational choice" strands of the other social sciences also use "methodological individualism" as the conceptual foundation for their theories and models. This means that social phenomena such as markets, political systems, crime, and culture can also be modeled as joint consequences of individual goals and constraints. Several examples of how such phenomena and their effects on markets can be analyzed using very similar models are developed in Part III.

Again, the claim is not that the rational-choice models are perfect, only that they shed useful light on a very broad range of social phenomena.

IX. Appendix to Chapter 2: Some Useful Mathematical Concepts and Definitions

A. **A function** is a mapping from one set (often Q or quantity in economics) into another set (such as net benefits, costs, benefits, utility, profits, revenue, etc.) Many of the mathematical properties of a given function can be deduced from its "shape."

B. One of the most widely used characterizations of a function's shape in economics is *concavity*. There are **three notions** of concavity used in economics, although in this course, only the first and second are used outside of this chapter.

- DEF: **Strictly Concave**: function f is *strictly* concave iff

$$\alpha f(X_1) + (1-\alpha)f(X_2) < f(\alpha X_1 + (1-\alpha)X_2) \quad \text{where } 0 < \alpha < 1.$$

Geometrically this means a function is strictly concave "if and only if" all the points on a line segment connecting any two points on a function always lies "beneath" the function of interest.

Strict Concavity is the assumption about functions that is most often used in this course.

A strictly concave function has at most one maximum, which allows us to characterize choices that are very specific—whether the choices are unconstrained or constrained by some convex set.

- DEF: Concavity: function f is concave if and only if (iff)

$$\alpha f(X_1) + (1-\alpha)f(X_2) \leq f(\alpha X_1 + (1-\alpha)X_2) \quad \text{where } 0 < \alpha < 1.$$

Geometrically this means that points on a cord (line segment) connecting any two points of a function always lies on or beneath the function of interest. Thus, a straight line is concave, but not strictly concave. Every strictly concave function is also concave, but not every concave function is strictly concave.

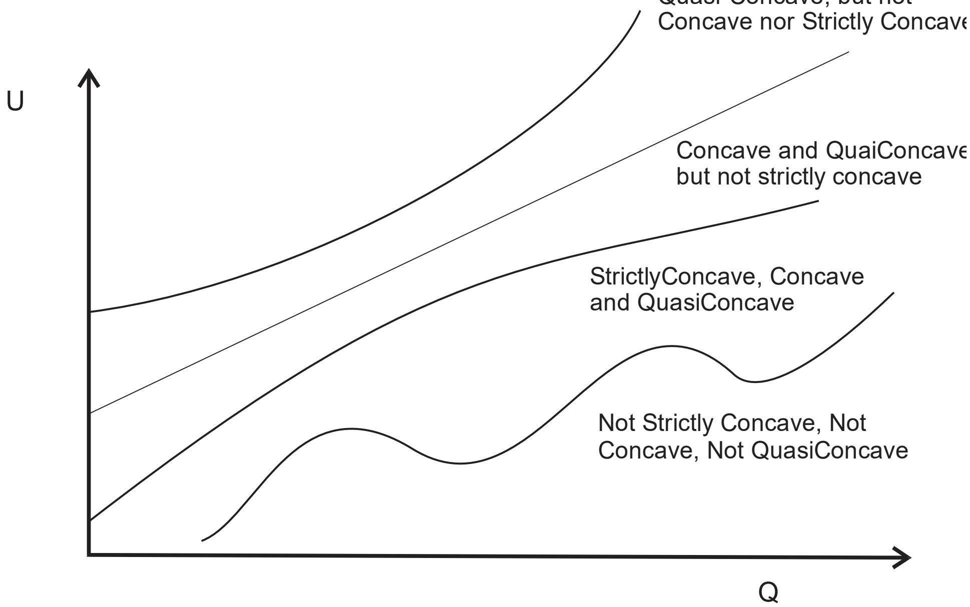
- DEF: Quasi-Concave: Concavity: function f is *quasi* concave iff

$$f(X_1) < f(\alpha X_1 + (1-\alpha)X_2) \quad where: f(X_1) < f(X_2)\ and\ 0 < \alpha < 1.$$

The values of a quasi-concave function always lies above the lower of the two end points of a cord connecting any two points on the function.

Any monotone increasing function is quasi-concave, but it is not necessarily concave or strictly concave, because it may increase at an increasing rate.

Figure 2.5: Concavity



Maxima and Minima of Functions

C. Strictly concave functions have a number of useful properties in the context of "optimizing" behavior.

- A *strictly* concave function has at most one maximum. (Draw some pictures to see why.)
- However, a **concave function** may have an infinite number of global maxima, but if there is more than one maximum, they make up a continuous linear interval. (A horizontal line is concave, but not strictly concave.)

D. DEF: The **global maximum** of a function, f(x), is a value, f(x*), that exceeds all others over the entire range of the function (e. g. for every neighborhood of x*).

E. DEF: A **local maximum** of a function, f(x), has a value which exceeds those of other points within a finite neighborhood of x*. That is, f(x*) is a local maximum if f(x*+e) < f(x*) and f(x*-e) < f(x*) for 0<e<E, for some E>0.

Note that if a function has a global maximum, then that global maximum is also a local maximum.

- However, because a function may have many local maxima, only one of those can be a global maximum.

Derivatives of functions can be used to characterize *sufficient* conditions for concavity, strict concavity, and therefore also for global maxima and minima.

- A function is *strictly concave* if its first derivative(s) is positive, and its second derivative(s) is negative over its entire domain.
- A function is concave if its first derivative is positive, and its second derivative is less than zero over its entire domain.

F. Functions may have local and global maxima, although most of the functions used in economic model-building are assumed to be strictly concave and so have at most one maximum (e.g. only one local maximum, which is also its global maximum).

- A function is at local maximum at point Q* if and only if (iff) its first derivative at Q* has the value zero and its second derivative is negative within a finite neighborhood around Q*.
- A point, Q*, is the global maximum of function f(Q) if its first derivative has the value zero at Q* and its second derivative is negative throughout the domain of the function. (Notice that in this case function f(Q) is strictly concave.)
- Maxima are, as it turns out, important for constrained optimization.

Within a particular domain, as with $0<Q<2$, any function, f(Q) will have a highest value.

This would be the constrained optima or maximum for function f(Q) within the domain from 0 to 2. It would be the "constrained" optimum.

Note that there may be more than one such optima, as when f(Q) is a horizontal straight line or a simple sine curve.

However, at least one maximum will always exist. This is simply a property of real numbers, within any set there will always be a largest value (number).

- When the function is strictly concave and the constraint is a convex set (as with the interval example above, $0<Q<2$) there will be a unique maximum (optimum).
- Thus, when a consumer has a constraint set that is convex (e.g. can choose any "bundle" within a particular convex set such as budget set), and attempts to maximize a strictly concave objective function such as a utility function, net benefit function, or profit function, the unique maximum will be the rational individual's choice.

Virtually all of the models of consumer choice are grounded in this very general property.

It is used, for example, to derive demand and supply curves from consumer and firm choices.

It is also used in Game theoretic settings where strategy sets are continuous rather than discrete.

G. DEF. A function is said to be **homogeneous of degree k**, if and only if whenever

$$Y = f(X) \text{ , then } f(\beta X) = \beta^k Y$$

- A production function that is homogeneous of degree 1 exhibits constant returns to scale. Doubling all inputs, exactly doubles output.
- Cobb-Douglas functions and linear functions through the origin (Y = ax) are homogeneous of degree 1.
- Occasionally, utility and production functions are assumed to be *homothetic*, a somewhat more general family of functions than homogeneous functions.

H. **DEF**. A **homothetic function** is a composite function of the form $H = h(Q(a, b))$ where Q is a homogeneous function and d$H/dQ > 0$ over the entire domain of h or $dH/DQ < 0$ of the entire domain of h. (E.g. a homothetic function is a monotone increasing or decreasing transformation of a homogeneous function.)

- Not all homothetic functions are homogeneous.
- Homothetic utility functions have linear income expansion paths.
- Similarly, homothetic production functions have linear output expansion paths. (*The slopes of the isoquants are the same along any straight line through the origin*.)
- Assumptions of homogeneity and homotheticity make models and their implication less general than they would have been with assumptions of monotonicity, concavity, and strict concavity, but the clarity of the results is often felt to warrant such assumptions.

References

Arrow, K. J. (1958). Utilities, attitudes, choices: A review note. *Econometrica: Journal of the Econometric Society*, 1-23.

Becker, G. S. (1976). *The Economic Approach to Human Behavior*. Chicago, IL: University of Chicago Press.

Bentham, J. (1793/1952). *Manual of Political Economy*, réédité in J. *Bentham (1952)*, *1*, 219-273.

Chiang, A. C. (1984*). Fundamental Methods of Mathematical Economics*. New York: Macmillan.

Debreu, G. (1959). *Theory of Value: An Axiomatic Analysis of Economic Equilibrium* (Vol. 17). Yale University Press.

Edgeworth, F. Y. (1881). *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*. London, UK: Kegan Paul.

Hicks, J. R., & Allen, R. G. D. (1934). A reconsideration of the theory of value. *Economica*, *1*(1), 52–76.

Jevons, W. S. (1871). *The Theory of Political Economy*. New York: Macmillan.

Marshall, A. (1890). *Principles of Economics*. London, UK: Macmillan.

Menger, C. (1871/2024). *Principles of Economics*, translated by J. *Dingwall and B.F. Hoselitz*. Auburn AL: Ludwig von Mises Institute.

Mill, J. S. (1848/1909). *Principles of Political Economy with some of their Applications to Social Philosophy*, (ed. William James Ashley) London: Longmans, Green and Co., 7th ed.

Pareto, V. (1906/1971). *Manual of Political Economy*, translated by A. S. Schwier. New York, NY: Augustus M. Kelley.

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

Samuelson, P. A. (1948). Consumption theory in terms of revealed preference. *Economica*, *15*(60), 243-253.

Samuelson, P. (1983) *Foundations of Economic Analysis*, extended edition. Cambridge MA: Harvard University Press.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*(1), 99–118.

Smith, A. (1776/1904). *An Inquiry into the Nature and Causes of the Wealth of Nations*, edited with an Introduction, Notes, Marginal Summary and an Enlarged Index by Edwin Cannan. London, UK: Methuen, 2 vols.

Stigler, G. J., & Becker, G. S. (1977). De gustibus non est disputandum. *The American Economic Review*, *67*(2), 76-90.

Walras, L. (1874/1954). *Elements of Pure Economics*, translated from the French by W. Jaffé. London UK: George Allen & Unwin.

Witt, U. (2001). Learning to consume–A theory of wants and the growth of demand. *Journal of Evolutionary Economics*, 11(1), 23–36.