# Chapter 3: Supply: Making Things to Sell

#### I. Introduction: Supply as Production for Sale

When economists talk about supply, they refer to activities undertaken to transform a group of materials into others of greater market value that "producers" plan to sell to others at a profit. Supply is not just making stuff or providing services, but making stuff and providing services that are to be sold to others. Although preexisting things may be sold—such as "used" cars and used books are often resold—for the most part the economic theory of supply involves selling newly produced things or selling more or less ephemeral services (by the hour) to others for money.

This is not to say that the term "supply" is never used in other ways. For example, that term may be used by economists to describe the food stuffs on the shelves of persons who raise and "can" some of their own fruit and vegetables without any intent to sell them. These and similar activities are often termed household production (Becker, 1965). In places and periods before commercial societies emerged, a good deal, perhaps most, of a family's time, knowledge, and materials were devoted to such productive activities. Foodstuffs were homegrown or the product of hunting in nearby forests, cloth was home spun, clothing was homemade, and cooking was conducted over fires fueled by wood harvested from nearby woodlands, and so forth. Although a lot of time was spent producing goods and services, relatively little of it was produced for sale—and thus it would be almost irrelevant for price theory. Much of what a household consumed in those days was self-supplied.

The difference between household production and production for sale in markets is largely a difference in objectives. In household production, the aim is to increase utility directly either because the production process itself is valued or the outputs are regarded to be more pleasing or less expensive than substitutes that could have been purchased in markets. Home producers allocate time among their various productive activities to maximize their utility from the outputs produced. Production for supply takes place with income or profits as the objective, which, in turn, is used by those receiving to advance the interests of persons selling the goods produced (artisans, proprietors, firm owners, etc.)—e.g. to increase their utility. So, maximizing utility is the ultimate aim of both types of productive activities, but production for sale does so indirectly though effects on personal or household income.

Intermediate cases also exist in which persons produce some things for themselves (as with vegetable gardens) and sell part of their produce in markets. Such cases would have been commonplace in the days when markets first arose. However, in well-developed commercial systems, production by economic organizations is entirely (or almost entirely) for sale to others and most people in a community, region, or country "hire themselves out for wages," which is to say that they seek paid employment in specialized forms so that they have the money to purchase goods and services produced by others or other specialized organizations. Dense and ubiquitous commercial networks emerged first in the nineteenth century in what these days is referred to as "the West." It is the market relationships of such commercial societies that neoclassical economics emerged to explain.

The first models of supply—and the ones that have attracted most of the attention of textbook authors—assume that specialized organizations called "firms" engage in the production and that firm owners are those who profit from selling the outputs produced. Because more income is better than less income (other things being equal), these models all assume that the aim of production is to maximize the net revenues (profits) that firms realize from selling the output(s) produced.

#### Economic Organizations, Costs, and Profits

In the days when village and town markets first emerged, producers and sellers were often single persons or small organizations that were largely staffed by family members. As somewhat larger organizations emerged they often remained family-based proprietorships, but with many nonfamily members employed by their firms. Until the mid-nineteenth century, such firms were often similar in size and similarly limited in their ability to increase their scale of operation. Thus, when demand expanded, additional supply would be provided by additional firms. This is the historical basis of the Marshallian theory of long run supply, a topic taken up towards the end of this chapter. Being relatively small organizations, it was relatively easy to monitor employees and assure that they remained "on task" during the working hours—which were often quite long during the 19<sup>th</sup> century.

The cost of producing the goods sold combined labor costs along with other inputs, such as iron for black smiths, wood for carpenters, and wool, linen, or cotton for spinners, and thread for clothiers, etcetera. In addition, there would be associated capital goods (tools of the trade) for each type of producer-seller. And, of course, an artisan's own labor would have an opportunity cost in household production and hiring themselves out for wages. In none of the professions listed, would

a firm or its employees be able to survive without the sale of their goods and services, and subsequent purchase of foodstuffs and housing from farms or other specialized suppliers of goods and services. Positive net revenues had to be realized to make their businesses feasible.

Even in cases in which most persons were farmers, village craftsmen and craftswomen had to rely upon markets to meet their daily needs. They sold goods and services to farmers and others, and the proceeds were largely used to purchase other goods from other venders. These characteristics were at least part of the rationale for Marshall's (1890) characterization of firms in very competitive markets, and it is that characterization that remains the foundation of neoclassical models of market supply.

Sellers in such markets have an interest in maximizing their net income from their commercial activities. This may be tempered a bit by other interests such as a reputation for quality or fairness in dealings with customers, but, holding those constant, profit-maximization is a reasonable first characterization of a firm-owner's interest in producing goods for sale. Their ultimate aims may be more complex than that, but most of their other aims are advantaged by increases in personal income.

#### II. The Geometry of Profit-Maximizing Production

Since net income is simply the revenues received from sales less the cost of producing the goods and services sold, it is a type of net benefit. Thus, the logic and geometry of the net-benefit maximizing models developed above for consumers also apply to a firm owner's decisions about supply, albeit with a few modifications. For example, in contrast to the consumer's choice, revenue is the benefit, rather than a cost, for firm owners. And, the firm owner's cost is determined by the process (technology) used to manufacture the goods or services to be sold and input prices, rather than the selling price of its output(s).

Figure 3.1 illustrates the proprietor's output decision. In the setting where firms can sell all that they want at a given market price (e.g. when firms are price takers), price is marginal revenue (and marginal benefit) and marginal cost is the cost in expenditures and the opportunity cost of proprietor efforts in making each successive unit of the product sold. (In cases in which the firm sells more than one good, it is assumed that such choices can be made for each product in a manner that is independent of one another). Notice that the diagram is very similar to the diagram of net-benefit maximizing choice in chapter 2. The profit maximizing output also occurs where marginal

benefit (here marginal revenue or price) equals marginal cost (here the cost of producing an additional unit of the good to be sold). Note that price is the marginal benefit for price-taking firms but the marginal cost for price taking consumers.



Figure 3.1: Maximizing Net Revenue / Profits

As in the consumer choice model grounded in net benefit maximization, profits (net revenues or net benefits) are maximized at the quantity where the marginal revenue curve intersects marginal cost curve. And, again this can be demonstrated by using areas under the marginal revenue and marginal cost cures to characterize net revenues. For example, the net revenues associated with output Q' is (a+b)-b=a, those with output Q\* are (a+b+c+d) - (b+d) = (a+c), and the net revenues (profits) at Q'' are (a+b+c+d+f) - (b+d+e+f) = (a+c - e). Note that the highest net revenue (a+c) occurs at Q\*, which is the output where MR=MC. Outputs below that level, such as Q', have profits that are below that associated with output Q\* (they lack area c), and output levels above Q\* have additional costs that exceed the revenues generated (area e). Given the basic geometry assumed (MR initially greater than MC, and MC upward sloping), any output to the left of Q\* will lack an area similar to c and any output to the right of Q\* will lose an area that resembles area e. Even if firms cannot easily measure marginal cost—if they truly maximize profits, they produce at the output level where MC=MR.

As before there are also cases where the best output is zero because revenues are always below the firm's costs—as when consumer willingness to pay for a product is below the average cost required to produce it. Thu, there are many goods that are never produced because no one would want to buy them at a price that would financially justify production. There are also diagrams in which it might appear that a proprietor or firm would want to produce infinite amounts of the good because the marginal revenue is always above the marginal cost of production—some computer software products resemble such cases—but ultimately wealth, income, and borrowing limits imply a bounded ability for consumers to pay for a good or service. Thus, such infinite supply cases are actually impossible unless there is a complete absence of scarcity, because consumers could not pay a positive price for all the units produced.

The geometry of deriving a firm's supply curve is also similar to that for a consumer's demand curve. When firms are price takers (e.g. adapt to a prevailing price) they will produce at their net revenue maximizing output levels. Derive a firm's supply curve in such cases requires one to vary price and plot the firm's output level for each price within the plausible range in which the firm might be selling its goods or services. In most derivations, its assumed that each firm produces a single good or, if it produces multiple goods, each production line is entirely independent of the others (which is not always true of course). Other prices—including input prices—are assumed to be constant as one traces out a supply curve. Figure 3.2 illustrates this process for the firm above.



Three prevailing market prices are assumed, and the proprietor's or firm's decision in each case is modelled as an effort to maximize net income or profit. In principle one does for every conceivable price, but in practice a few prices are tried, and the rest of the curve is "smoothly" interpolated.

In the case depicted, marginal cost is assumed to fall at first, as the firm tools up for producing the good or service for sale, but after some point (here fairly early on) marginal costs begin to rise because of diminishing returns in producing the good or service of interest. This "nonmonitonicity" case has implications for the profit maximizing decision of the firm, because there are some outputs where MR=MC, but which yield a negative profit because costs are greater than revenues. Such outputs are never produced (except my mistake).

 $P_3$  has been chose to make area I equal to area II in the lefthand diagram. For this firm, output  $Q_3^*$  is the "break-even" or "shutdown" output for price  $P_3$ . It is the smallest output that this firm will ever produce, and  $P_3$  is the lowest price at which the firm will keep its doors open. At any price lower than  $P_3$ , the firm realizes a negative profit for every possible positive output, and so it is not able to sustain itself. At the ( $Q_3^*$ ,  $P_3$ ) price and output combination, the area of loss (area a) exactly equals the area of profit (area b). (Recall that the area under the MR curve from zero to quantity  $Q_3$  is total revenue, and the area under the MC curve is total cost of production or total variable cost, if fixed costs exist. Given the assumption that area I equal area II, profits are zero when the prevailing price is  $P_3$ .)

The supply curve consists of the profit-maximizing outputs for all prices at which profits are greater than zero (or more precisely, greater than the proprietor's opportunity cost rate of return). The supply curve includes the subset of the points on the firm's marginal cost curve at which profits greater than zero are realized (or more precisely, where positive profits are expected to be realized). Had the marginal cost curve been monotonically increasing, the firm's supply curve would have included all the points on the firm's marginal cost curve for  $Q \ge 0$ . In such cases, the supply curve goes through all of the same points as the firm's marginal cost curve, rather than a subset of them as in figure 3.2.

In the case where MC is monotone increasing in output, the MC and firm supply curves are inverse functions for one another. The two functions go through all the same points, but the direction of the mappings differs. In the case illustrated by figure 3.2, the points in common for each curve represent the subset of the output-cost space in which the supply curve and marginal cost function are inverse functions. Marginal cost maps from outputs (Q) into dollars per unit costs (P). Supply maps from prices (dollars per unit) into output levels. The two functions, thus, map in opposite directions. In cases, like the one depicted in figure 3.2, some points on the marginal cost curves (those at which profits are never maximized) do not appear as points on the supply curve. However, every point on the supply curve is also a point on the marginal cost curve. The neoclassical theory of the firm often makes a distinction between **long run and short run** supply. In the short run, there are fixed costs that cannot be avoided or changed in a reasonable short period of time. Fixed costs affect profit levels but not output levels in the short run. In the long run, all factors of production can be varied and so there are no fixed costs. This difference implies that the short and long run marginal cost curve differ, with a firm's long run marginal cost curve being "flatter" or more "price elastic" than the short-run marginal cost curve because the proprietor or firm's management has more flexibility in choosing production methods in the long run than in the short run and so can do more to keep marginal costs down. (Usually, it assumed that it easier to vary labor than capital, so capital is assumed to be fixed in the short run but not in the long run, although this is not always the case. Some forms of human capital are quite costly and time consuming to acquire.)

However, the essential geometry of a price-taking firm's output decisions is not affected. It chooses an output that sets marginal revenue equal to marginal cost. If no outproduces profits greater or equal to zero, then it will shut down and produce no output. In a diagram, the only difference between long run (lr) and short run (sr) marginal cost and supply is their labels . One should label the long run MC curve as MC<sup>LR</sup> the short run MC curve as MC<sup>SR</sup>. If both are drawn on a single diagram the MC<sup>SR</sup> curve rises more steeply than the MC<sup>LR</sup> curve, because fewer adjustments can be made to reduce production costs in the short run than in the long run.

Characterizing exceptions to the "rule of thumb" that profit maximizing firms "all" produce outputs where MR=MC is generally easier with geometry than with calculus. (One has to use a method referred to as the Kuhn-Tucker method to deal with those cases using calculus.) So, most calculus-based models assume that the MC is monotone increasing in output levels, even though it is clear that few if any firms ever produce infinitesimal levels of output.<sup>1</sup>

## III. Deriving Supply Curves from Cost Functions Using Calculus

In Chapter 2, the heart of deriving individual demand function was a consumer's total benefit or utility function. The heart of deriving a firm's supply curve is its total cost function. A total cost function maps possible output levels into dollar costs or costs measured with some other currency. A cost function includes all the variables that affect costs. Among these are output levels,

<sup>&</sup>lt;sup>1</sup> Mathematically characterizing such "corner solutions" is normally done using the Kuhn-Tucker approach, which is beyond the scope of this textbook.

input prices, technology of production, regulations, taxes on inputs or waste products, and so forth. However, for the purposes of deriving a supply curve, all of the determinants of cost except output levels are normally held constant. This assumption does not imply that the other factors are unimportant. They affect both the shape and extent of supply through their effects on marginal cost. Rather it is because a supply function characterizes the relationship between prevailing market prices and a firm's output level—holding other factors constant. As cost determinants other than output level change, so will the location of the supply function or its geometric representation as a supply curve. As developed below, supply functions include all those neglected variables as arguments. However, for the neoclassical theory of price determination, it is the relationship between market prices and output levels that is the primary focus of attention.

Holding all the other determinants of a firm's production costs constant, allows relatively simple functional forms to be used to characterize a firm's cost function—with the effects of the other factors implicitly determining constants or exponents in the cost function used.

As, in Chapter 2, relatively simple functions will be used initially and more general and abstract functions afterwards.

## **Deriving Supply Functions from Exponential Functions**

First, consider a simple concrete functional form for a cost function:

$$C = a + bQ^c \tag{3.1}$$

with C begin total cost, Q being the proprietor or firm's output level, and a, b, and c being parameters of the exponential cost function. Well assume that a and b are greater than zero and that c is greater than one. These assumptions assure that costs are greater than zero and that marginal cost rises with output levels reflecting diminishing marginal returns in production. They will also assure that the profit function is strictly concave. Constant "a" is the firm's fixed costs, which exist only in short run models of cost. Coefficient b can be thought of as the cost of the inputs required to increase output and exponent c as an indicator for the rate at which the required inputs increase as output increases. Each term in this simple cost equation, thus, has an implicit economic meaning or implication.

Revenue for a proprietor or firm selling goods in a market in which he or she is a "price taker" is simply R = PQ, where R is revenue, P is the prevailing market price, and Q is his or her output level. His, her, or its profit,  $\Pi$ , is simply total revenue less total cost:

$$\Pi = PQ - (a + bQ^c) \tag{3.2}$$

The output that maximizes profit, Q\*, can be found by differentiating the profit function with respect to quantity and setting the result equal to zero.

$$\frac{d\Pi}{dQ} = P - bcQ^{c-1} = 0 \tag{3.3}$$

The first term (P) is marginal revenue. Total revenue increases by amount P every time another unit is produced and sold. The second term ( $bcQ^{c-1}$ ) is marginal cost. It is the rate at which total cost increases as output increases. Thus, equation 3.3 implies that a firm owner that is interested in maximizing profits should choose an output where marginal revenue (P) equals marginal cost (( $bcQ^{c-1}$ ), as in the geometric case examined earlier in the chapter. Note that fixed costs (a) play no role in short run output decisions whenever net revenues greater than zero (ignoring fixed costs) are realized by the firm, as in this case.

To be sure that this "first order" condition characterizes a maximum rather than a minimum, the second derivative of 3.2 should also be calculated and its "sign" determined (e.g. whether it is always greater than zero, less than zero, or may have different "signs" as Q varies). This is possible for most concrete function forms of profit functions. In this case, the second derivative is:

$$\frac{d\Pi^2}{dQ^2} = -bc(c-1)Q^{c-2} < 0 \tag{3.4}$$

Notice that b, c, and (c-1) are all greater than zero so the slope of the marginal cost curve is positive as is consistent with diminishing marginal returns. However, there is a minus in front of that slope characterization, so the entire term is negative, and it is so for any quantity greater than zero. Thus, the profit function is strictly concave in the domain of interest (e.g. where Q>0). In choice settings in which firms are price takers, the concavity of the profit function is entirely determined by the shape of the cost function. (Marginal revenue in the "price taker" case is simply the price at which its output is sold. When MR=P, the marginal revenue function (P) has zero slope, and thus it does not affect the magnitude of the second derivative.)

To characterize the supply curve, we solve equation 3.3 for Q as a function of market price. To do so, first shift all the terms with Q in them (here just one) to the lefthand side of the equation by adding  $bcQ^{c-1}$  to each side.

$$bcQ^{c-1} = P \tag{3.5}$$

Divide each side by *bc*, and then raise each side to the  $\frac{1}{c-1}$  power. These steps generate:

$$Q^* = \left(\frac{P}{bc}\right)^{1/c-1} \tag{3.6}$$

This is the **supply function** for the firm's output decision being modelled.

Note that the amount produced and sold increases with price. This family of total cost functions implies that supply curves always slope upward (e.g. output is monotone increasing in selling price). Price is in the numerator. Thus, every supply function from this family of exponential cost functions has an important property that economic intuition suggests a supply function should have—they are upward sloping—and conversely, the results also show that economic intuition about a firm's supply decisions is consistent with at least one fairly general family of cost functions.

There are other factors that also implicitly affect output as. For example, holding price constant, the quantity produced for sale falls as "c" increases, which is to say, as the rate at which the additional inputs required to increase output increases. It also falls as the average cost of inputs (b) increase. Factors that affect the marginal cost of production affect output decisions.

These results are also consistent with economic intuitions about factors that cause a supply curve to shift. Again, economic intuitions are affirmed by a logically consistent model of decision making—although it is possible that they are supported by only a subset of possible cost functions.

## Deriving Supply Functions from More General Cost Functions

More abstract and encompassing families of cost functions can also be used to characterize a firm's supply function. For example, suppose that the firm's cost function is simply:

$$C = c(Q, w, r) \tag{3.7}$$

with C begin total cost, Q being the firm's output level, *w* being the prevailing wage rate for labor and *r* being the cost of capital. Assume that the first derivatives of the cost function are all positive and the second derivatives are also all positive (to reflect diminishing marginal returns in production). The cross partials are also assumed to be positive. Total cost naturally rises as output increases and input prices, other things being equal. The assumptions about second derivatives assures that cost function is convex and that the profit function is strictly concave. The cavity of the profit function in the Q dimension assures that there is at most one unique output at which profits are maximized for a given price. Chapter 4 will demonstrate why input prices belong in every firm's cost function—although it should be intuitively obvious for most economics majors.

Revenue for firms that are "price takers" is simply R = PQ, where R is revenue, P is the prevailing market price, and Q is his or her output level. Its net revenue (profit) is denoted as  $\Pi$  and is simply total revenue less total cost:

$$\Pi = PQ - c(Q, w, r) \tag{3.8}$$

The output that maximizes profit, Q\*, can be found by differentiating the profit function with respect to quantity and setting the result equal to zero.

$$\frac{d\Pi}{dQ} = P - dC/dQ = 0 \quad \text{at } Q^* \tag{3.9}$$

The first term (P) is marginal revenue. Total revenue increases by amount P every time another unit is produced and sold. The second term (dC/dQ) is marginal cost. It is the rate at which total cost increases as output increases. Thus, equation 3.9 implies that a firm's owner profits should choose the output where marginal revenue (P) equals marginal cost ((dC/dQ), as in the geometric case examined earlier in the chapter, if he, she, or they want to maximize profits.

To be sure that this "first order" condition characterizes a maximum rather than a minimum, the second derivative of 3.2 should also be calculated and its "sign" determined (e.g. whether it is always greater than zero, less than zero, or may have different "signs" as Q varies). This is possible for most concrete function forms of profit functions. In this case, the second derivative is:

$$\frac{d\Pi^2}{dQ^2} = -\frac{dC^2}{dQ^2} < 0 \tag{3.10}$$

So, the profit function is strictly concave in the  $\Pi x Q$  plane, given the assumption that the marginal cost function is upward sloping. (As in the previous geometric model, marginal revenue in the "price taker" case is flat (MR=P), and thus, the marginal revenue function has zero slope and does not affect the magnitude of the second derivative of profits with respect to quantity.)

The marginal cost function,  $\frac{dc}{dq}$ , implicitly includes the same argument as its "parent" function, which is to say, it is a function of Q, w, and r, just as the total cost function is. The conventional notation hides this, but when dealing with abstract functions, it has to be kept in mind, because it affects the calculations that can be undertaken with such functions.

To characterize the supply curve, we make use of the implicit function theorem. Recall that any (locally) differentiable function that has the value zero at all relevant points, such as  $h(P, Q^*, w, r) = 0$ , has the property that each variable can be described as a function of every other. Note that equation 3.9 has this form for Q\*, the firm's profit maximizing output. Thus, we can characterize Q\* as

$$Q^* = s(P, w, r) \tag{3.11}$$

Equation 3.11 is the firm's supply curve and the letter given to the function characterized was chosen with this in mind; "s" is intended to remind the reader that this is a supply function.

Note that the amount produced and sold varies with price, and in this case, with the price of inputs. How those variables affect supply is of interest.

To determine the slope of the supply curve, recall that the implicit function differentiation

rule states that if  $H \equiv h(P, Q^*, w, r) = P - \frac{dC}{dQ} = 0$ , then  $\frac{dQ^*}{dP} = \frac{\frac{dH}{dP}}{-\frac{dH}{dQ^*}}$ . In this case that is a relatively simple expression:

$$\frac{dQ^*}{dP} = \frac{1}{-\left(-\frac{dC^2}{dQ^2}\right)} > 0 \tag{3.12}$$

The derivative of equation 3.9 with respect to P is just 1. (The cost function does not include the price of the output to be sold.). The derivative of equation 3.9 with respect to Q is also quite straightforward. It is simply the second derivative of the profit function which is equation 3.10 above. From that, we know that  $\frac{dH}{dQ} < 0$ , and thus  $-\frac{dH}{dQ} > 0$ . Thus, supply functions for this quite general family of cost functions all slope upwards. Both the numerator and denominator of  $\frac{\frac{dH}{dP}}{-\frac{dH}{dQ}*}$ are greater than zero.

Again, we find that economic intuitions grounded in the geometry of supply decisions are not "wrong" or logically impossible—in fact, they are actually quite general. They apply to a very broad family of possible cost function—every cost function with an upward sloping marginal cost curve.

#### IV. Market Supply by Price Taking Firms in the Short Run and Long Run

To derive a **short run market supply** function from individual firm supply functions, one simply adds up the supply curves (the quantities sold at a particular price) of every firm that is in the market at time t, N<sub>t</sub>, in which case short run supply is:  $S_t = \sum_{1}^{N_t} s_{it}(P, w, r)$ , where  $s_{it}$  is the i<sup>th</sup> firm's short run supply curve at time t. If the firms are identical (as assumed in the Marshallian approach), short run market supply is simply N<sub>t</sub> times the typical firm's short run supply curve,  $S = N_t s(P, w, r)$ .

There are two different models of long-run supply in competitive markets—both of which have their merits, the Marshallian model and the Ricardian Model. Their main difference is that the Marshallian model assumes that every firm has the same cost function, and the Ricardian model assumes that firms may have different cost functions. Whether long run cost functions differ or not is likely to vary by market and also may vary through time.

During the late nineteenth century and early twentieth century, the majority of markets were served by relatively small family-based enterprises. These were mainly organized as proprietorships and partnerships. These businesses had limited ability to grow because of the manner in which they were organized. There was often a single proprietor or small number of active partners who made production decisions and monitored their firm's employees. Efficient firms were of limited size. Partly because of this, Alfred Marshall who wrote the most widely used neoclassical economics textbook, believed that most long run supply adjustments took place entirely through the entry and exit of similarly sized firms. At the long run equilibrium, each firm sold its output at the lowest possible price (similar to the output in Figure 3.2 that produced zero profits). Firms would exit (often through bankruptcy) when the price fell below this level, and new firms would enter when it rose above that price. (Net revenue did not literally fall to zero, but rates of profit fell to the "normal" rate of return, which many economists refer to as zero profits—e.g. zero excess profits.)

Because producers were pretty similar in the manner in which the produced the services that they sold, they also tended to be of roughly the same size. Thus, as a first approximation, one could think of every firm as being identical. In that case, the equilibrium number of firms in the industry at a given price was simply market demand at that price (the lowest selling price) divided by the size of the least cost level of output (that which minimized long run average costs). For example, if the marginal cost function illustrated in figure 3.2 is a typical firm's long-run marginal cost, then all firms

would be of size  $Q_3^*$  in the equilibrium that emerges in perfectly competitive Marshallian market, and the number of firms in the industry would reflect market demand at price  $P_3$ .

In the long run, Marshall argued—and most other neoclassical textbook models have used his assumptions and reached similar conclusions—that each firm's profits would be zero or, more precisely, be equal to the typical proprietor's or partnership's opportunity cost rate of return on their investments, labor, and entrepreneurship. (These entrepreneurial and investment costs are rarely explicitly modeled, but they should be considered part of a firm's long run total cost function. They can be approximated by "r" the rental cost of capital owned or rented by the firm owner.)

In today's world, non-chain carpenters, plumbers, mechanics, coffee shops, bicycle shops, restaurants, and clothing stores still resemble the Marshallian market environment. Firms in these markets are all roughly the same size and use very similar production and/or selling methods. Entry and exit are constant, and the average firm earns a "normal" rate of return (although some individual firms appear to earn above normal rates of return). When demand increases, existing firms expand a bit, and other small-scale entrepreneurs enter to realize profits associated with the new higher demand for the products being produced and sold. When times are tough, many firms exit through bankruptcy.

An alternative to the Marshallian model, which this author prefers, is the Ricardian model. In a **Ricardian competitive market**, firms have somewhat different production functions, and therefore a somewhat different cost and supply functions. Profits vary among firms in such markets because of differences in their cost functions. Only the marginal firm earns no profit (e.g. realizes the "normal" opportunity cost rate of return). The others earn different levels of profit depending on their cost functions. Obvious examples of such markets include many natural resource markets because some deposits of minerals (as with oil) are easier to harvest than others. Similarly, a piece of farmland may be closer to market, have more fertile land, be in a region with more fortuitous weather, in a region with lower wage rates, etc. Also, an individual firm may have an especially talented workforce or entrepreneur that can keep costs down and quality up.

In cases in which entry takes a good deal of time or is very costly, one can ignore entry and exit and simply focus on each existing firm's long-run supply curve. Market supply in such cases (assuming that there are enough firms, so that the price taking assumption is reasonable) is simply the sum of the long run supply curves of the firms in the market.  $S = \sum_{i=1}^{N} s_i(P, w, r)$ . Profits will

vary among firms and only the last or smallest firm will earn zero profits (e.g. be indifferent between continuing in the market of interest or moving to another product or regional market).

For Ricardian markets the area-based diagrams of the first part of the chapter can be applied for both long run and short run analysis. For the Marshallian case, long and short run diagrams differ, with long run supply typically being a horizontal line with each firm producing at the minimum of its long-run average cost curve (at an average cost of  $P_3$  and scale equal to  $Q_3$  in figure 3.2).

#### V. The Geometry of Supply by Firms with Price Setting Abilities (Monopoly Pricing)

Firms are not always "totally" constrained by prevailing market prices—which is to say by the ready availability of very good or perfect substitutes for their products. In some places, there may be only a single firm that produces a particular product or service. For example, in moderately sized towns, there may be only a single coffee or bicycle shop, or a single restaurant the provides meals based on cuisines from other places. In the United States, for example, there are many towns where there is only a single restaurant that sells meals from the French, Indian, Chinese, or Mexican cuisines. Such firms compete with other firms for the purchases of local users of restaurants, but they are not entirely constrained by that competition to sell their output at a particular prevailing market price—because there is no prevailing market price for their "unique" product or service.

In larger markets, patent and copyright protection may assure that some products are produced by only a single producer, author, or advisor. Or, an innovative firm may produce products that no one can easily copy because of trade secrets. Such merchants each have their own downward sloping demand curves for their products.

Such firms have some "monopoly power," the degree of which varies with the extent to which their general type of product or service is demanded and the ease with which their own particular service or product can be substituted for by other readily available goods or services. For example, a restaurant goer in a small town might have a Mexican meal rather than a French one, even though he or she prefers French to Mexican cuisine, because the French restaurant charges "too much" for its unique menu of meals. Even when products are not identical, substitutions are often possible.

Neoclassical models of monopoly pricing imply that such firms will make output decisions that take account of how their decisions affect the selling price of their goods and thus their total

revenues and profits. Figure 3.3 illustrates the output decision of such a firm. The fact that it faces a downward sloping demand curve—no matter how sharply downward sloping it is—implies that its marginal revenues are no longer a determined by a horizontal line equal to price. (The calculus illustration in the next sub section will make it clearer why that is the case.)





Given the results from chapter 2 and the assumption that firms are profit maximizers, we know that this firm (as others) will choose an output where the marginal cost of production equals the marginal revenue generated by it. The product will be priced so that it all sells, so the price is that at which demand will exactly equal supply, which is the one labeled P\*.

The demand curve has been assumed to be "linear" (e. g. a straight line), which it turns out implies that the marginal revenue is also a straight line. In particular it is a line that lies halfway between the demand curve and the horizontal axis (for reasons that the calculus below will make clear). The area that corresponds to the firm's profits can be calculated in two ways. First, one can use the "area" approach developed in Chapter 2. The firm's total revenue at Q\* is the area under the marginal revenue (MR) curve from 0 to Q\*. Its total (variable) cost is the area under its marginal cost curve from 0 to Q\*, and its profits are the difference between the two areas. In the diagram, it is the somewhat irregular shaped triangle between the marginal cost (MC) curve and the marginal revenue curve.

Another area that corresponds to total revenue is simply the rectangle whose area is  $P^*Q^*$ , which turns out to be the rectangle characterized by the point ( $P^*$ ,  $Q^*$ ) on the demand curve and the horizontal and vertical axis. Using this measure for total revenue generates another area for profits which is the somewhat irregular area between the marginal cost curve and the horizontal line equal to  $P^*$ . Since the two profit measures have to be the same (the firm only earns one profit, a real number) it turns out that the triangles labeled I and II have to have exactly the same areas. (In fact, they are identical congruent triangles given the linear demand and MR curves.) This geometric regularity exists because of the two ways that total revenue can be calculated, and it does not have important economic implications, although it shows that one need not focus entirely on marginal revenue and marginal cost to characterize profits.

The main point here is that some products have only less than "perfect" substitutes and therefore producers of such goods have some ability to set their own prices. Those prices are still constrained by demand, and profit maximizing outputs are still determined marginal costs and marginal revenues associated with their choice settings. However, the marginal revenue curves in such settings tend to be downward sloping rather than horizontal lines equal to the prevailing market prices for their goods or services.

## VI. The Calculus of Output Choices by Firms with a Bit or More of Monopoly Power

To illustrate the mathematical model and calculus that lies behind Figure 3.3, consider the case where a firm faces a downward sloping demand curve Q = a - bP and has a total cost function  $C=zQ^s$  where z>0 and exponent "s" is greater than or equal to 1.

To characterize a firm's total revenues in terms of output, we'll need to first find the inverse of the demand function, which is to say the function describes the prices at which output Q can be sold at—given the firm's demand curve. This requires solving Q = a-bP for price as a function of output. Subtracting "a" from both sides of the equation and dividing by "-b" generates the equation P = (Q-a)/(-b) = (a-Q)/b, which describes the price at which output Q can be sold. Total revenue is just PQ as before. However, in this case P is affected by the firm's output decision.

Using the equation that we just derived which describes the price at which Q units can be sold, we can now characterize total revenue as a function of the firm's output:  $PQ = [(a-Q)/b]Q = (aQ-Q^2)/b$ . Next, we write down the profit function that takes account of the effect of the firm's output on revenues and costs.

$$\Pi = TR - TC = (aQ - Q^2)/b - zQ^s$$
(3.13)

To find the profit maximizing output, differentiate the profit function with respect to Q and set the result equal to zero. The "first order condition" characterizes the profit maximizing output.

$$\frac{a-2Q}{b} - szQ^{s-1} = 0$$
 (3.14)

Notice that the first term is marginal revenue. It characterizes the rate at which total revenues increase as output increases. Note that it is linear falls at twice the rate that the linear demand curve did, at 2Q/b rather than Q/b, which accounts for the fact that the MR curve in figure 3.3 was halfway between the demand curve and the horizontal axis. The second term is marginal production cost. It characterizes how total cost increase as output increases. The profit-maximizing output occurs at the output, Q\*, where MR = MC, as in figure 3.2.

To make sure that the first order condition characterizes a maximum rather than a minimum or inflection point, we take the second derivative of the profit function. If it is always less than zero (negative) when Q>0, then the profit function is strictly concave in the relevant domain. Thus, if there is a  $Q^*$  that satisfies the first order condition, that output level will be at the profit maximizing output. The second derivative is the derivative of equation 3.14 which is:

$$-2b - (s - 1)szQ^{s - 2} < 0 \tag{3.15}$$

Which is less than zero for all quantities greater than zero. "b" is positive, so the first term is negative. (s-1), s, and z are positive, so the second term is also negative (because of the leading negative sign). So, this profit function characterized by equation 3.13 is strictly concave, and the solution to equation 3.14 to characterizes the firm's output. The price at which the output is sold is the price from the demand curve, which can be determined by substituting Q\* into the inverse demand curve, P\* =  $(a-Q^*)/b$ .

With concrete functional forms, one can often—but not always—solve the first order condition (here, equation 3.14) for an equation that characterizes Q\* in terms of the exogenous parameters of the firm's choice setting. In this case it is not possible to do so for the general class of exponential cost functions used to this point.

On the other hand, special cases from the assumed family of cost functions can be solved. To illustrate such a case, suppose that exponent s=1, in other words that the firm's cost function has constant marginal costs. In this case equation 3.14 takes the form:

$$(a - 2Q^*)/b - z = 0$$
 (3.14 b)

And, a bit of algebra allows  $Q^*$  to be characterized as a function of parameters of the firm's choice setting. First, add  $2Q^*/b$  to both sides to obtain:

$$a/b-z = 2Q^*/b$$

Next, reverse the sides of the equality, and multiply both sides by b/2 to get:

$$Q^* = (b/2) [a/b-z] = \frac{a}{2} - (bz)/2$$
 (3.15)

Equation 3.15 characterized the price-making firms output in terms of parameters of the demand function (here a and b) and of its cost function, (here z). By inspection, we can see that the higher "a" is the greater is output. However, the greater are b and z, the lower is the profit maximizing output, other things being equal. If we knew precise values for a, b, and z, we would know the firm's precise output. Its price can be determined by substituting Q\* into the inverse demand equation worked out at the beginning of this section. In the special case used to obtain equation

3.15, P\* = (a-Q\*)/b, or, substituting for Q\*, P\* = 
$$\frac{a - \left[\frac{a}{2} - \frac{bz}{2}\right]}{b} = \frac{a(z-1)}{b} - z/2.$$

## An Abstract Model of Output Choices by a Price Making Firm

Even more abstract version of a "price making" firm's decision can be characterized. For example, we can assume that a firm faces the demand curve, Q = q(P,Y,N,Pj) where P is the price of the firm's product, Y is average consumer income, N is the number of consumers in this market, and Pj is the price of a substitute or complement for the firm's product. The firm's cost function can be characterized as C=c(Q, w, r) where Q is the firm's output and w is the wage rate paid for its labor and r is the rental cost for firm's capital equipment. Demand functions similar to this one were worked out in Chapter 2. Cost functions similar to this one are worked out in the next chapter. Profits are again simply total revenue less total costs, or  $\Pi=PQ-C$ .

To characterize the profits as a function of output, we'll again need to find the inverse demand curve. Notices that a "zero function" can be created by subtracting the demand function from the desired output by consumers, Q - q(P,Y,N,Pj)=0. That allows us to use the implicit function theorem to describe sales price as a function of output, P=p(Q, Y,N,Pj). Using that function, the profit function becomes:

$$\Pi = \mathbf{p}(\mathbf{Q}, \mathbf{Y}, \mathbf{N}, \mathbf{P}\mathbf{j})\mathbf{Q} - \mathbf{c}(\mathbf{Q}, \mathbf{w}, \mathbf{r})$$
(3.16)

To simplify a bit, it will simply be assumed that the profit function is strictly concave. Given this the first order condition will characterize the firm's profit maximizing output. The first order condition is simply:

$$\frac{d\Pi}{dQ} = \left(\frac{dP}{dQ}\right)Q + P - \frac{dC}{dQ} = 0$$
(3.17)

The first two terms are the firm's marginal revenue function. (Notice that we've used the product differentiation rule to find the firm's marginal revenue function.) The last term is its marginal cost function. So again, the firm's profit maximizing output, Q\*, is where marginal revenue equals marginal cost.

At Q\*, the first derivative of the profit function equals zero and so the implicit function theorem can again be applied to characterize any variable of interest in terms of the others in the revenue and cost function being assumed. We are, of course, most interested in Q\* and we'll want to write down the function that describes it.

$$Q^* = s(Y, N, P_j, w, r)$$
(3.18)

The selling price is just  $P^*=p(Q^*, Y, N, Pj)$ . The firm's output and selling price both vary with parameters of the demand curve for its product (here, Y, N, Pj) and its cost function (w and r).

The implicit function differentiation rule can be used to characterize the effects of changes in the exogenous variables on the firm's output. Recall that if we define equation 3.17 as H, the partial derivative of Q\* with respect to wage rates, w, is  $H_w/-H_Q$  where subscripts are being used to denote partial derivatives with respect to the variable subscripted. Because the profit function has been assumed to be strictly concave,  $H_Q$  will be negative, and  $-H_Q$  will be positive, so the sign of the derivative is determined by the numerator.

The signs of the partial derivatives in the numerator are easy to characterize for the cost variables. For example,  $H_w$  is simply - dc<sup>2</sup>/dQdw which is less than zero because an increase in wage rates increases marginal costs (dc<sup>2</sup>/dQdw >0). Increases in marginal costs induce decreases in output and increases in prices (because the demand curve slopes downward).

The derivatives with respect to the determinants of demand are a bit more complicated because we have used the implicit function theorem to characterized prices as a function of outputs, thus the implicit function differentiation rule would have to applied both the "P" terms in the f.o.c using the zero equation that we used to characterize the price-output relationship. That relationship would affect the derivatives of both terms with P in them. For example, if G = [Q - q(P,Y,N,Pj)], where G is the zero equation used to obtain the inverse demand function, then  $dP/dY = G_Y/-G_P =$  $-q_Y/-(-q_P)$ , both terms being partial derivatives of the demand function, q. The term dP/dQ would equal  $G_Q/-G_P$ , or  $G_Q^*(-G_P)^{-1}$  and its derivative with respect to Y would be require applying the division or multiplication rule as with:  $dP/dQY = G_{QY}^*(-G_P)^{-1} - G_Q^* G_{PY} (-G_P)^{-2}$ .

For the purposes of this chapter, the main conclusions that emerge from a more general derivation of a firm's supply decision is that the supply decisions of firms with a bit of monopoly power are determined by jointly by the factors that determine their cost functions (as also true of price taking firms) and also of factors that determine the extent of their demand. These include such considerations as average consumer income, number of consumers, and the prices of substitutes and/or complements to the firm's output.

Although these choices are more complex than those of price-taking firms, they do provide a clear explanation for how prices are set and adjusted in such markets. In markets where firms have no monopoly power, the explanations for prices and price adjustments usually rely upon unmodelled minor adjustments in prices by firms or an invisible Walrasian auctioneer or price-driven inventory adjustments. These later explanations for equilibrium prices in competitive markets are taken up in Chapter 5.

## References

- Alchian, A. S. and Allen, W. R. (1974). University Economics, Elements of Inquiry (third edition). Saddle Hill, NJ: Prentice Hall.
- Becker G. S. (1965). A Theory of the Allocation of Time, Economic Journal 75: 493-517.
- Marshall, A. (1890). Principles of Economics. New York: Macmillan and Company.

Muellbauer, J. (1974). Household Production Theory, Quality, and the" Hedonic Technique". *The American Economic Review*, *64*(6), 977-994.