Chapter 3: Supply: Making Things to Sell

I. Introduction: Household Production versus 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 those undertaking the transformation (the producers) plan to sell to others. Supply is not just making stuff or providing services, but making stuff and providing services that are *sold* to others in in their communities or extended trading networks. Although preexisting things may be sold—such as "used" cars and 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 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 activities of persons who raise some their own vegetables in gardens, bake bread, and construct sheds to store tools in their backyards, without any intent to sell them. Such activities are often termed household production (Becker, xxxx). 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. Much of what a household consumed in those days was self-supplied. 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.

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. Production for supply takes place with income or profits as the objective, which, in turn, is used by those receiving to advance their interests—utility. So, maximizing utility is the aim for each type of production, but production for sale does so indirectly though effects on personal or household income.

Intermediate cases exist in which persons produce some things for themselves (as with vegetable gardens) and sell part of their produce in markets. However, in well-developed commercial systems, production by economic organizations is entirely for sale to others and most people in a community, region, or country "hire themselves out for wages" that they use to purchase goods and services. 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.

II. The Geometry of Supply Decisions by "Price Taking" Firms

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 were developed in the nineteenth century, they often remained family-based proprietorships, albeit with many more non-family members employed by their firms. Such firms tended to be 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 long run supply, a topic taken up in the next 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 19th century.

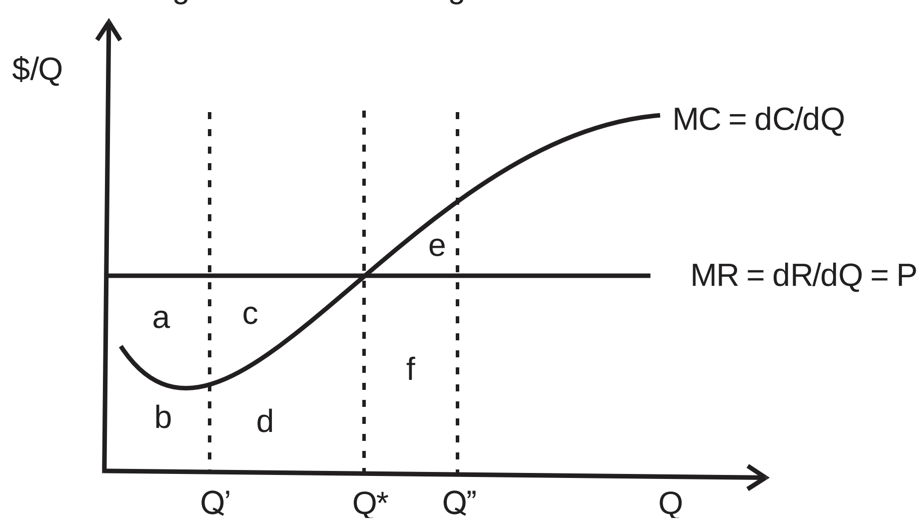
The cost of producing the goods sold combined labor costs along with other inputs, such as iron for black smiths, wood for carpenters, and wool, linen, or cotton for spinners, and thread for clothiers, etcetera. In addition, there would be associated capital goods (tools of the trade) for each type of producer-seller. In none of the professions listed, would a firm or its employees be able to survive without selling their goods and purchasing foodstuffs from farms or markets for produce. Even in cases in which most persons were farmers, such village craftsmen and craftswomen had to rely upon markets to meet their daily needs. They sold goods and services that 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 (xxxx) characterization of firms in very competitive markets, and it that characterization that remains the foundation of most 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 and fairness in dealings with customers, but, holding those constant, profit-maximization is a reasonable first characterization of a proprietor'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.

Since net income is simply the revenues received less the cost of production, it is a form of net benefit and the logic and geometry of net-benefit maximizing models developed above for consumers will apply to a firm owner's decisions about supply. However, in contrast with the consumer's choice, revenue is the benefit, rather than the cost, and cost is determined by the process (technology) used to manufacture the goods or services to be sold, rather than prices paid for the final goods.

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, 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). The profit maximizing output occurs where marginal revenue (here price) equal marginal cost (here the cost of producing one more unit of the good to be sold).

Figure 3.1: Maximizing Net Revenue / Profits

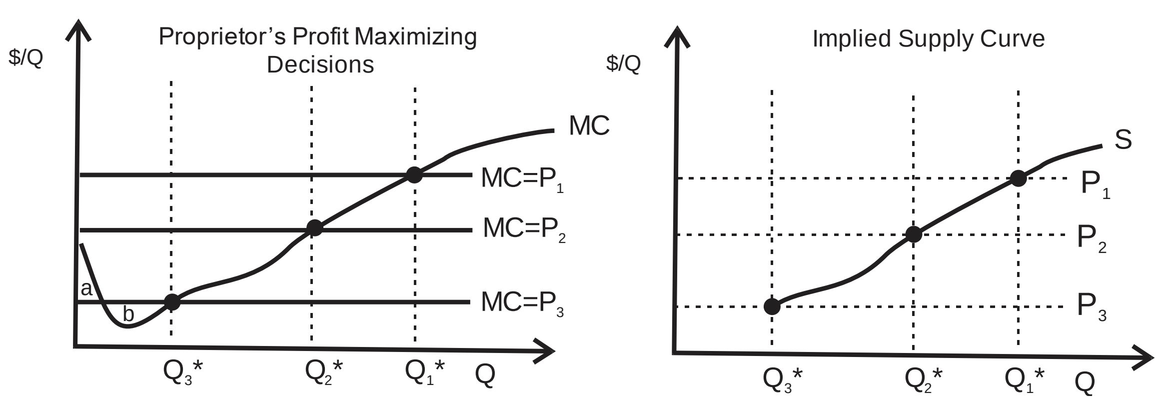


As in the consumer choice model grounded in net benefit maximization. It is the intersection of the marginal revenue (the proprietor's marginal benefit) and marginal cost curve that determine the ideal output. And, again this can be demonstrated by using areas under the marginal revenue and marginal cost cures to characterized the net revenues at Q' , (a+b)-b=a, at Q*, (a+b+c+d) – (b+d) = (a+c), and at Q", (a+b+c+d+f) – (b+d+e+f) =(a+c - e). The highest net revenue (a+c) occurs at Q*. Outputs below that level such as Q' have profits that are below that associated with output Q* (area c less), and output levels above Q* have additional costs that exceed the revenues generated (area e). (Note that 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.)

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 produces 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 may 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.

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. Figure 3.2 illustrates this process for the firm above.

Figure 3.2: The Geometry of Supply



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 are "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 "non-monitonicity" assumption 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).

For this firm, output Q_3* is the "break-even" or "shut-down" output for price P_3. It is the smallest output that this firm will ever rationally 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 will earn a negative profit on every possible positive output, and so is not be 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.)

The supply curve consists of 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 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 $Q \geq 0$. In this case, the supply curve goes through all of the same points as the firm's marginal cost curve.

However, as in the case of the geometric derivation of demand curves, where the MB and consumer demand curves were inverse functions, the marginal cost curve of a firm and its supply curve 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 do, however, go through the same points. 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 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 amount of time. Such 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 not 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.

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, or equal to zero if profits greater or equal to zero cannot be realized (ignoring fixed costs in the short run). The only difference is that one should label the MC curve as the long run curve, as with MC^{LR}.

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.[1]

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, including such things as output levels, 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

[1] Mathematically characterizing such "corner solutions" is normally done using the Kuhn-Tucker approach, which is beyond the scope of this textbook.

level change, so will the location of the supply function or its geometric representation as a supply curve. As seen in the next chapter, 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. Assume that a and b are greater than zero and that c is greater than one. These assumptions, as will be seen, 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 roughly 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 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) plays 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:

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

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 always is so for any quantity greater than zero. Thus, the profit function is strictly concave. In this case, the concavity of the profit function is entirely determined by the shape of the cost function in settings where firm owners sell their product in markets at "given" prices. (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.)

To characterize the supply curve, we solve equation 3.3 for Q as a function of market price. A few algebraic steps can 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.

$$bcQ^{c-1} = P \qquad (3.5)$$

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

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

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

Note that the amount produced and sold increases with price. (Price is in the numerator.) and falls as "c" increases, which is to say, as the rate at which additional inputs are required increase. It also falls as the cost of inputs (b) increase. Thus, every supply function from this family of exponential cost functions has the property that economic intuition suggests a supply function should have—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. Thus, economic intuitions are not inherently "wrong" or logically impossible—although it is possible that they are reliable only for 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, assume 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). The cross partials are also assumed to be positive. Total cost, of course, rises as output increases or if input prices increase. The assumptions about second derivatives will help to assure that cost function is convex and the profit function is strictly concave. The cavity of the profit function in the Q dimension is most important for the derivation of the slop of a firm's supply function, as demonstrated below. Chapter 4 will demonstrate why input prices belong in a cost function—although it should be intuitively obvious to most undergraduate economics majors.

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:

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

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 owner that interested in maximizing profits should choose the output where marginal revenue (P) equals marginal cost ((dC/dQ), as in the geometric case examined earlier in the chapter.

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:

$$\frac{d\Pi^2}{dQ^2} = -\frac{dC^2}{dQ^2} + \frac{dC}{dQdr} + \frac{dC}{dQdr} < 0 \qquad (3.10)$$

This way of writing the second derivative of the cost function with respect to quantity takes account of the possible effects of changes in output on the mix of inputs employed and thereby how input prices affect marginal production costs. In this case, we have already assumed that $\frac{dC^2}{dQ^2}$ and the two cross partials are greater than zero, and thus it follows that $-\frac{dC^2}{dQ^2} + \frac{dC}{dQdr} + \frac{dC}{dQdr} < 0.$[2] So, the profit function is strictly concave in the ΠxQ plane, given the assumption that the marginal cost function is upward sloping. (As in the previous 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.)

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 was. The conventional notation hides this, but when ever dealing with abstract functions, it has to be kept in mind, because it affects a variety of calculations that can undertaken with such functions.

---

[2] Sometimes this second derivative will be written as simply $\frac{dC^2}{dQ^2}$. This is either being used as a shorthand for the longer expression, or implicitly assuming that the cross partials are all zero. Notice if the wage rate and interest rates terms were left out of the cost function, $\frac{dC^2}{dQ^2}$.would be all that is necessary.

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, naturally "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≡ 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:

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

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 more complicated because of the possible effects of output on the input mix, an possibility that is demonstrated mathematically in the next chapter.

Again, we find that economic intuitions grounded in the geometry of supply decisions are are not "wrong" or logically impossible—and actually quite general insofar as they apply to a very broad family of possible cost function.

IV. Market Supply with Price Taking Firms

To derive a short run market supply function, one adds up the supply curves of every firm that is presently in the market. To derive a long run supply function one adds up the long run supply curves of every firm that would be in the market at a given price (after entry and exit takes place). In some cases, this is simply the number of firms already in the market, in other cases adjustments in the number of firms in the industry is a major factor in the long run. Indeed, in the Marshallian approach to long run supply entry and exit are the main adjustment mechanism.

During the late nineteenth century and early twentieth century, which was the period in which neoclassical economics was worked out—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, and there were limits in the number of employees that they could organize, supervise, and monitor. Thus, efficient firms were of limited size and Marshall believed that most long run supply adjustments in such markets took place through the entry and exit of such firms. Firms would exit when the stopped producing the product of interest and took up producing others or became bankrupt. Firms would enter when the left some other market or a proprietor or small group of partners organized a new business.

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 them as being the same size, and market supply at a given price or level of demand was simply "N" the number of efficient sized firms necessary to meet demand at that price. In the long run, Marshall argued, and most other neoclassical models of perfect competition through to the present agreed, that profits would be zero or equal to the typical proprietor or partnerships opportunity cost rate of return. In today's world, non-chain coffee shops, bicycle shops, restaurants, and clothing stores resemble the Marshallian market environment. They are all roughly the same size. Entry and exit are easy in the long run, and as a consequence, the average firm earns a "normal" rate of return. When demand increases, such firms expand a bit, and many other small-scale entrepreneurs enter to meet the new demand, driving down prices and profits. When times are tough, many firms exit, reducing short run losses and decreasing supply. The short run supply curve at a given moment is simply N_t (the number of firms that exist at a given moment) times the typical firm's short run supply curve, $S = N_t\, s(P, w, r)$.

An alternative to the Marshallian model, which this author prefers, is the Ricardian model. In a market with a Ricardian long run supply curve, every firm has a somewhat different production function and therefore a somewhat different cost function and supply function. Profits vary widely among firms in such markets because of differences in their cost functions. Only the marginal firm earns no profit or simply the opportunity cost rate of return. The others earn different levels of profit depending on their costs. Obvious examples of such markets include many natural resources—and oil deposit may be easier or harder to exploit, a peace of farmland may be closer to market, have more fertile land or be in a region with more fortuitous weather, a firm may have an especially talented workforce or entrepreneur that can keep costs down and quality up. In such cases, one can essentially ignore entry and exit and simply focus on each 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. $S = \Sigma s_i(P, w, r)$. Profits will vary among firms and only the last or smallest firm will earn zero profits (e.g. have net revenues equal to its opportunity cost rate of return).

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, because in the absence of technological externalities (that affect production costs) long run supply are always a horizontal line with each firm producing at the minimum of its long-run average cost curve (a point similar to Q_3 in figure 3.2.

V. Supply By a Price Making Firm

Essentially all mathematical and geometric models of the behavior of firms and consumers in competitive markets assume that both consumers and firms are "price takers." That is to say that they assume that market prices are exogenously determined and that consumers and firms simply adjust to those prices. Consumers do so by purchasing their utility or net-benefit maximizing quantity of goods. Firms do so by producing their profit maximizing quantities of goods (given technology and input prices) given the prevailing price of the product(s) that they sell. One way to bring price determination into economics models is to drop the assumption that firms are price takers that sell homogeneous products. If firms in an industry produce slightly different products or are in different locations, each firm faces a downward sloping demand curve. And given that demand curve, they would naturally choose a price and output combination that maximizes their profits—other things being equal.

Firms that compete with other firms selling similar but not identical products would have relatively horizontal, "flat," or price-elastic demand curves. Such markets are sometimes termed "monopolistically competitive" markets. Prices in such markets would tend to be fairly similar because their products are similar. Prices would be fairly close to their marginal costs (as in perfectly competitive markets) and profits would be modest and mostly be within a fairly narrow range (as in perfectly competitive markets). These are the firms that neoclassical theory of monopolistic competition attempts to model. The more similar the products and better informed the consumers in such markets are, the less firms are able to charge prices above their marginal costs. Such markets can be regarded as "very competitive" rather than perfectly competitive.

In contrast, firms that dominate a market without good substitutes (or serious rivals producing identical products) are called monopolists. They usually face a "steeper" downward sloping (less-elastic) demand curve. They are said to have significant "monopoly power," because they can set their prices well above the marginal cost of production. In such cases, if the market is relatively large, their profits would tend to be well above average and their prices would be significantly greater than marginal cost. These are the firms that the neoclassical theories of monopoly attempts to model.

It turns that that very similar mathematical models can be used to characterize the pricing and output decision of both kinds of firms, because what distinguishes them from firms in perfectly competitive markets is simply that they face (and know that they face) downward sloping demand curves. To illustrate the mathematics, we'll go through an example using concrete functional forms. Then we'll draw a diagram of that solution, and finally redo the solution in somewhat more abstract form.

Characterizing Monopoly Output Decisions Using Concrete Functions

Suppose the demand curve faced is linear and downward sloping $Q^D = a - bP$ and the firm's cost curve is upward sloping $C = cQ^2$. The firm's profit is

$$\Pi = PQ - C \tag{3.13}$$

We want to express price, P, in terms of the quantity produced and brought to market, Q, so that the effects of the firm's output on market prices can be taken into account when it makes its output decision. This can be found by solving the demand curve for P as a function of Q. The demand function, $Q = a - bP$, implies that:

$$P = (Q - a)/(-b) = a/b - Q/b \tag{3.14}$$

This result and the cost function can be substituted into the profit function to characterized profit in terms of its output decision—its supply:

$$\Pi = PQ - C = (a/b - Q/b)Q - cQ^2 \tag{3.15}$$

Differentiating with respect to Q and setting the result equal to zero yields:

$$d\,\Pi/dQ = [a/b - 2Q/b] - 2cQ = 0 \quad \text{at } Q^* \tag{3.16}$$

The first term (the one in brackets) is marginal revenue, and the last term is marginal cost.

Solving this expression for Q characterizes this firm's ideal output level (which we denote as Q*). One method of solving for Q is the following. First, shift all of the Q terms to the lefthand side.

$$a/b = 2Q/b + 2cQ$$

Factor the righthand side, then then solve for Q.

$$a/b = Q\,(2c + 2/b)$$

$$Q^* = (a/b)/(2c + 2/b)$$

This expression can be simplified a bit by multiplying the top and bottom of the fraction (the one characterized by the terms in the two parentheses) by (b/b) to get:

$$Q^* = a/(2bc + 2) \tag{3.17}$$

The equilibrium price is found by substituting this quantity (or expression for quantity) into the pricing equation derived above from the firm's demand function. Recall that, $P = a/b - Q/b$. Substituting our result for Q* into this function yields.

$$P^* = a/b - Q^*/b \quad = a/b - [a/(2+2bc)]/b \tag{3.18}$$

This expression (the one to the right) can be simplified by bringing "b" inside the brackets.

$$P^* = a/b - a/b(2+2bc)$$

Then placing both terms over the same denominator b(2+2bc)

$$P^* = a(2+2bc)/b(2+2bc) - a/b(2+2bc)$$

Then combining both terms

$$P^* = (2a+2bca - a)/b(2+2bc)$$

Subtracting the "a" from the "2a" and simplifying

$P^* = (a+2bca)/b(2+2bc) = a/2b(1+bc) + 2bca/2b(1+bc)$ which simplifies to

$$\mathbf{P^* = a/2b(1+bc) + ac/(1+bc)} \tag{3.19}$$

The marginal cost at Q^* is simply the first derivative of the cost function evaluated at Q^*. In this case, the marginal cost is is $dC/dQ = 2cQ$. The value of marginal cost at Q^* is

$$MC(Q^*) = 2c(Q^*) = 2c[a/(2+2bc)] = 2ca/2(1+bc)$$

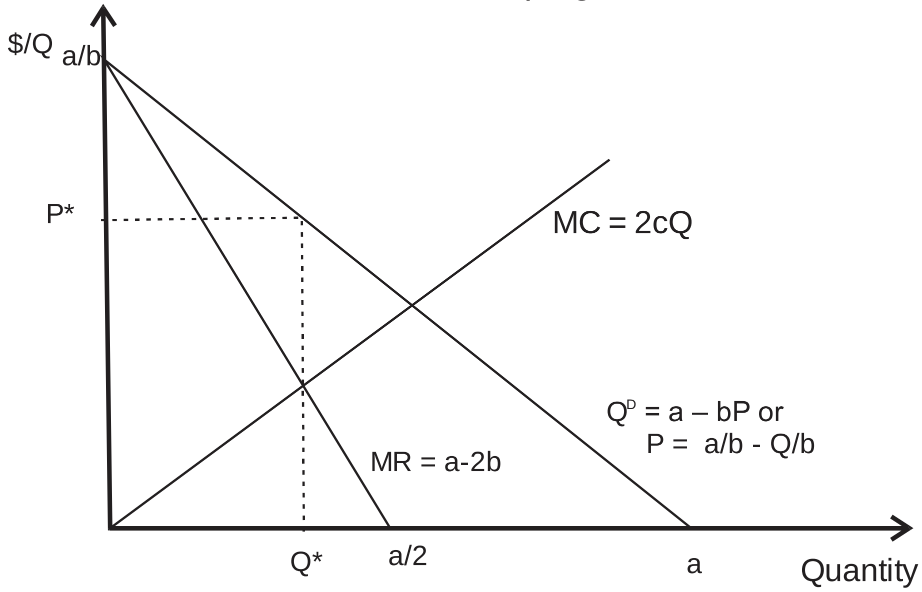
Which simplifies to $$MC(Q^*) = ac/(1+bc) \qquad (3.20)$$

Recall that this was the second term in the expression for P^* derived above (written in blue).

The algebra implies that the price charged for a monopolist's output, P^*, is greater than its marginal cost at Q^* by amount $a/2b(1+bc)$. Note that this monopolistic "mark up" declines as the downward slope ($-1/b$) of the demand curve (here, in its inverse form, with P on the left) increases. Written in this way, which is the usual way of thinking about the slope of a function on a diagram, as "b" increases, the demand curve gets flatter. It becomes more and more horizontal, most likely because of the availability of better substitutes.

Figure 3.3 below illustrates this monopolistic firm's output and pricing decision. The original demand curve (with Q on the lefthand side) is a linear expression with an intercept on the horizontal axis at "a," the value of Q when P is zero. When expressed with P on the lefthand side ($a/b - Q/b$) the vertical intercept is a/b, the value of P when Q is zero.) Marginal revenue is: $MR = a/b – 2Q/b$ (the first two terms in the first order condition for profit). Note that MR also has the value a/b when $Q = 0$. And MR is itself zero when $Q=a/2$. It emerges from the vertical axis at the same point as the linear demand curve and falls twice as fast. The first order condition derived above is satisfied when Q is such that MR = MC. The price at which the goods are sold is that on the demand curve at Q^*.

Figure 3.3: Supply when a Firm Faces a Downward Sloping Demand Curve



Characterizing Monopoly Output Decisions Using Abstract Functions

A more general derivation of the supply and pricing decisions of firms facing downward sloping demand curves can also be undertaken. The assumptions and mathematics parallels the concrete functional form derivation, but in this case we will not develop a particular function that describes the firms supply decision, but rather a general function. It will, however, have derivatives that can be characterized and in some cases signed.

To begin with assume that the firm faces a demand function such as $Q = q(P, Y)$ where P the price of the good to be supplied and Y is average consumer income. Assume that the firm has a cost function of the form $C=c(Q, w, r, T)$ and for the purposes of this derivation assume that it is a strictly convex function. To characterized profit in terms of quantity, we'll need to characterized the demand curve as a function of Q rather than of P. The simplist way to do this is to note that the demand function implies that $0 = Q - q(P, Y)$, which is a differentiable "zero function" and so the implicit function theorem can be used to characterize price, P, as a function of sales, Q, as with:

$$P=f(Q, Y) \tag{3.21}$$

The firm's profits can thus be written as:

$$\Pi = f(Q, Y)Q - c(Q, w, r, T) \tag{3.22}$$

Differentiating with respect to Q and setting the result equal to zero generates that first order condition that characterized the firms output Q*.

$$\frac{d\Pi}{\mathrm{dQ}} = \left(\frac{df}{dq}\right) Q + f(Q, Y) - \frac{dC}{dQ} = 0 \quad at\ Q^* \tag{3.23}$$

This is another zero function and the implicit function theorem implies that Q* can be characterized as a function of the other variables in the functions in the first order condition. In this case, that yields:

$$Q^* = g(Y, w, r, T) \tag{3.24}$$

Q* is the supply function of the price-taking firm. Its output varies with consumer income, input prices, and production technology. The price at which that output will be sold is determined by equation 3.21.

$$P^* = f(Q^*, Y) \tag{3.25}$$

Derivatives of both equations can be developed using the implicit function differentiation rule. In the case of income, it would have to be used twice, once for the demand function (equation 3.21) and once for the first order condition. Derivatives of the other variables would require just one application. Note for example that the derivative of the firm's output with respect to the wage rate that it has to pay is:

$$\frac{dQ^*}{dw} = (-d^2C/dQdw)\ /\ (-\frac{d^2\Pi}{\mathrm{dQ}^2}) \quad < 0$$

First note that the denominator is just the second derivative of the profit function, which is negative is the profit function is strictly concave (at least "near" Q*). The negative of a negative is positive, so the sign of the denominator is positive. Thus, the sign of this derivative is determined by the numerator. The numerator is the negative of the effect of an increase in wage rates on marginal cost. Since an increase in wage rates increases marginal cost, the cross partial, $\frac{d^2C}{dQdw}$, is positive and the negative of that cross partial is negative. So the quantity supply by a price making firm tends to fall as its input prices increase. Thus, the effect on supply is qualitatively similar to that for competitive or price taking firms.

VI. A Few Brief Conclusions about Supply

The analysis of supply decisions undertaken in this chapter demonstrates that such decision are fundamentally driven by cost considerations in both very competitive and somewhat less competitive or noncompetitive settings. This follows regardless of whether firms are price takers or price makers. It also turns out that the effects of input prices on those decisions are qualitatively the same for firms that operate in all three settings. Higher input prices tend to reduce supply in each type of market, whereas lower input prices tend to increase it.

In the next chapter, we'll explore how input prices affect decision about how to produce the things and services brought to market and provide somewhat deeper foundations for the theory of supply developed in this chapter by providing somewhat deeper foundations for the cost functions assumed.