9 A Second Look at Firm Behavior Under Perfect Competition

This chapter covers ...

- how profit-maximizing firms behave in competitive markets (behavioral foundation of the supply function).
- how the supply function is related to marginal and average cost functions and what this says about the informational demands and effective organization of firms.
- the technological prerequisites for the functioning of competitive markets.
- how competition drives profits to zero and why this is not bad.

9.1 Introduction

> The natural price or the price of free competition ... is the lowest which can be taken. [It] is the lowest which the sellers can commonly afford to take, and at the same time continue their business. (Adam Smith, The Wealth of Nations (1776[1991]), Book I, Chapter VII)

This chapter will take a closer look at the supply decision of a firm that sells in a market with perfect competition. To make the problem manageable, one has to specify the objective of the firm and say a few words about its ownership structure as well as its internal organization.

The standard assumption in the literature is that firms seek to *maximize profits*. ✍ If p is the price of some good produced by the firm, y is the quantity produced and $C(y)$ are the costs of production, then the profits are $\pi(y) = p \cdot y - C(y) = R(y) - C(y)$, where $R(y)$ stands for revenues. One way to think about this objective function is in terms of the interests of the owners. Assume that a single person owns the firm and uses it as a vehicle to maximize her income. What objective would she try to give the firm in order to pursue her goal? Obviously, the increase in income that the owner can extract from the firm is equal to the firm's profit: the owner deploys capital and labor, which costs her $C(y)$ for y units of output, and

M. Kolmar, *Principles of Microeconomics*, Springer Texts in Business and Economics,
DOI 10.1007/978-3-319-57589-6_9

she gets the revenues of the firm, $p \cdot y$. Therefore, the surplus or the increase in income is equal to the firm's profit. Hence, if income-maximizing owners invest in firms, it is in their best interest that the firms maximize profits. If owners invest in firms because they want to reach something else, then the imputed objective may be different. Nevertheless, it is a good starting point to conjecture that most shareholders invest in corporations because they want to make money.

$\mathcal{B}$

Digression 22. The Limits of Profit Maximization: Information, Contracts, and the Organization of Firms

The idea that income-maximizing owners would like to make sure that the managers of the firm maximize profits is simple and powerful. However, it is the source of a lot of controversy for both normative and positive reasons. From the positive point of view, it is sometimes argued that firms do not, in fact, maximize their profits. Deviations from this objective may have several reasons. They can be a result of imperfect information about costs and revenues. Limited information is definitely a relevant problem and it may lead to decisions that are apparently not in line with profit maximization. Nevertheless, it does not falsify the objective *per se*. As previous chapters have shown, it is the purpose of managerial accounting to provide information to support decisions. If the information is bad, the decisions are bad, and the first impulse should be to develop a better accounting system, not to abandon profit maximization.

Another important reason for deviations from profit maximization results from the fact that firms are usually complex networks of individuals with their own objectives. The key question then becomes whether it is possible to align the interests of the owners with the interests of the workers. Take the CEO of a firm that is not managed by the owner and assume further that both, the owner and the manager, want to maximize their incomes. The income of the manager depends on the contract, so it becomes a problem of contract design whether the owner's and the manager's income maximizations coincide. (One can think of such a contract as an incentive mechanism. An optimal contract is one that creates no externalities between manager and owner.) The key question is, therefore, how such a contract has to be designed to make sure that the manager internalizes the interests of the owner. If the contract is ill-designed, the manager will use her discretionary power to maximize her own income, which is not compatible with the owner's income maximization and, therefore, is in conflict with profit maximization. An example might be a contract with bonus payments that incentivizes short-term profits, despite the fact that they are in conflict with the long-term interests of the firm.

To simplify things, one assumes that owners want to maximize income, that contracts perfectly align owners' and managers' interests, and that the accounting system is sufficiently precise to allow for a realistic view of costs and revenues.

Hence, firms maximize profits. This case acts like a benchmark. If one understands the benchmark, one can get a better understanding of the effects of deviations from it.

It makes sense to state the objective function explicitly. The firm maximizes profits by the choice of the quantity of the good produced. Formally, it is expressed as follows:

$$\max_{y} p \cdot y - C(y).$$

The assumption of perfect competition enters the above choice problem, because the price is treated as a parameter, which means that the firm takes it as given. The above formulation also assumes both that the quantity produced and the quantity sold is identical and that firms do not produce for or sell from stocks.

The concept of marginal revenues will be helpful in understanding optimal firm decisions.

► **Definition 9.1, Marginal revenues** The marginal revenue of production is the revenue the firm makes by an additional unit of production: $MR(y) = dR(y)/dy$.

What are the implications of the assumption of profit maximization for the supply decision of the firm? The following thought experiment allows one to gain a better understanding. Assume that the firm produces a quantity such that marginal revenue is larger than marginal cost and marginal costs are strictly positive. What would the effect on profits of an increase in production by one unit be? Given that profit is revenues minus costs, profit must increase, if marginal revenues exceed marginal costs, so it would be rational to increase production, because the firm would make more money with the additional unit than it would cost. Next, assume that the firm produces a quantity such that the marginal revenues are smaller than the marginal costs. In this case, profit would go down, because the next unit of production costs more than the firm would get for it on the market. Therefore, it would be rational to reduce production. These two observations pin down the profit-maximizing behavior of a firm: the optimal quantity is the one where marginal revenues are equal to marginal costs.

This condition can also be derived analytically, by setting the first derivative of the profit function equal to zero. Given that the price is fixed for a firm on a competitive market (the firm is a price-taker), the marginal revenues are equal to the price of the good, $MR(y) = p$, and one gets:

$$\pi'(y) = p - C'(y) = p - MC(y) = 0.$$

Denote the quantity of the good that fulfills this condition by y^*. This result is a very important finding in the theory of firm behavior: a profit-maximizing firm on a competitive market produces according to the "price-equals-marginal-costs" rule, because marginal revenue is equal to the price under perfect competition. This rule has several implications that the following paragraphs will discuss. Its applicability also depends on several factors that one has to make explicit for an in-depth understanding of its role in the theory of firm behavior. I will start with the implications.

The first and most important implication, for an economist, is the link between the cost and supply functions. The condition $p = MC(y)$ formally establishes a relationship between price p and quantity y. The supply function of a firm establishes the same type of relationship with $y = y(p)$. It maps each price onto a quantity produced by the firm. If one looks at the inverse of the marginal-cost function, one gets $y = MC^{-1}(p)$. But this mapping has prices as domain and quantities as codomain. This mapping described by the supply function has quantities as its domain and prices as its codomain. The implication of this is that these two mappings are inverse to each other and that a competitive firm's supply function is identical to its inverse marginal-cost function. When one observes a firm's market behavior, one can "look through" the supply decision and get information about the firm's marginal costs.

This finding also allows for a more in-depth understanding of the willingness-to-sell concept, which Chap. 5 introduced: I have argued that one can interpret a point along the supply function as the minimum price the producer has to get in order to be willing to sell an additional unit of the good. This price, as we have seen, is equal to the marginal costs of producing this unit, which makes perfect sense: marginal costs measure how much it costs to produce an additional unit of the good. If one is paid more than that, one makes a profit with this unit, if one is paid less, one takes a loss. Therefore, one is indifferent between selling and not selling, if one gets exactly one's marginal costs.

$\mathcal{B}$ The "price-equals-marginal-costs" rule also has important managerial implications: In order to be able to behave in accordance with this rule, a manager needs information about the market price and the marginal-cost function. This has implications for the organization of her company: in addition to the factory that produces the goods that the firm sells, the firm needs a controlling department that collects information about costs as well as current (and maybe also expected future) market prices. The organization of a competitive firm is not very complicated. However, getting the controlling correct is crucial for the firm, because the quality of decisions depends on the accuracy of the information about marginal costs.

Unfortunately, life on competitive markets, as either a manager or an economist, is not as simple as the above rule suggests. Next, one has to put the "price-equals-marginal-costs" rule into perspective. I will discuss three aspects in the following subchapter: technological conditions under which perfect competition works, short-versus long-run decisions, and the relationship between firm and market supply.

9.2 The Relationship Between Production Technology and Market Structure

As this chapter has already shown, the profit-maximizing production decision of a firm can be characterized by the condition:

$$\pi'(y^*) = p - C'(y^*) = p - MC(y^*) = 0.$$

I will scrutinize this approach from a purely technical point of view and discuss the economic implications thereafter. This approach illustrates how a back-and-forth between economic thinking and mathematical reasoning can improve one's understanding of the economy.

One may remember, from one's mathematics classes in high school, that the so-called first-order conditions are necessary, but not sufficient, for the characterization of a maximum. The only thing that a first derivative of zero guarantees is that the function has a "flat" point, which can be a maximum, a minimum or a point of inflection. In order to make sure that one characterizes a maximum, one has to check the so-called second-order conditions, i.e. one has to check if

$$\pi''(y) = -C''(y)$$

fulfills a certain property at the potential optimum. The second-order condition says something about the curvature of the function. In order to make sure that the first-order condition characterizes a maximum, one has to make sure that the profit function is "hump shaped" or, more technically, strictly concave. (In addition, one has to make sure that there is an interior optimum. A technical condition that guarantees this is $p > MC(0)$ and $p < \lim_{y\to\infty} MC(y)$.) This is guaranteed if the second derivative of the profit function is negative,

$$\pi''(y) = -C''(y) < 0 \Leftrightarrow C''(y) > 0.$$

Figure 9.1 illustrates this case.

The upper part of Fig. 9.1 shows the profit of a firm for all possible production levels. It is inversely u-shaped. The first-order condition identifies the quantity where the slope of the function is zero, which characterizes the maximum profit. The lower part of Fig. 9.1 disentangles the profit into revenues and costs. The straight line represents revenues as a function of y. Revenues are linear in production, because it is simply the product of an exogenous price and the endogenous quantity of the good. Costs increase disproportionally. Profit in Fig. 9.1 is equal to the *vertical* difference between the revenue and the cost curve. What the firm tries to do is to identify the output where this vertical difference is at its maximum. This point is at the place where both functions have the same slope. The slope of the revenue function is p and the slope of the cost function is $MC(y)$.

In concluding that purely technical argument, is there anything one can learn from this condition as an economist? First of all, one can see that the condition restricts the class of admissible cost functions to those that increase overproportionally in production. Hence, the marginal costs of production increase in the quantity produced. Here is an example: assume that one invests time to study for the final exam. The more time one spends, the better one's expected grade becomes. It is relatively easy to pass, but it becomes more and more difficult to get the best possible grade. This property is nicely reflected by an application of the so-called 80-20 rule (also called the Pareto principle), which states that, for many events, 80% of the effects come from 20% of the causes. Applied to the example, it would say that one

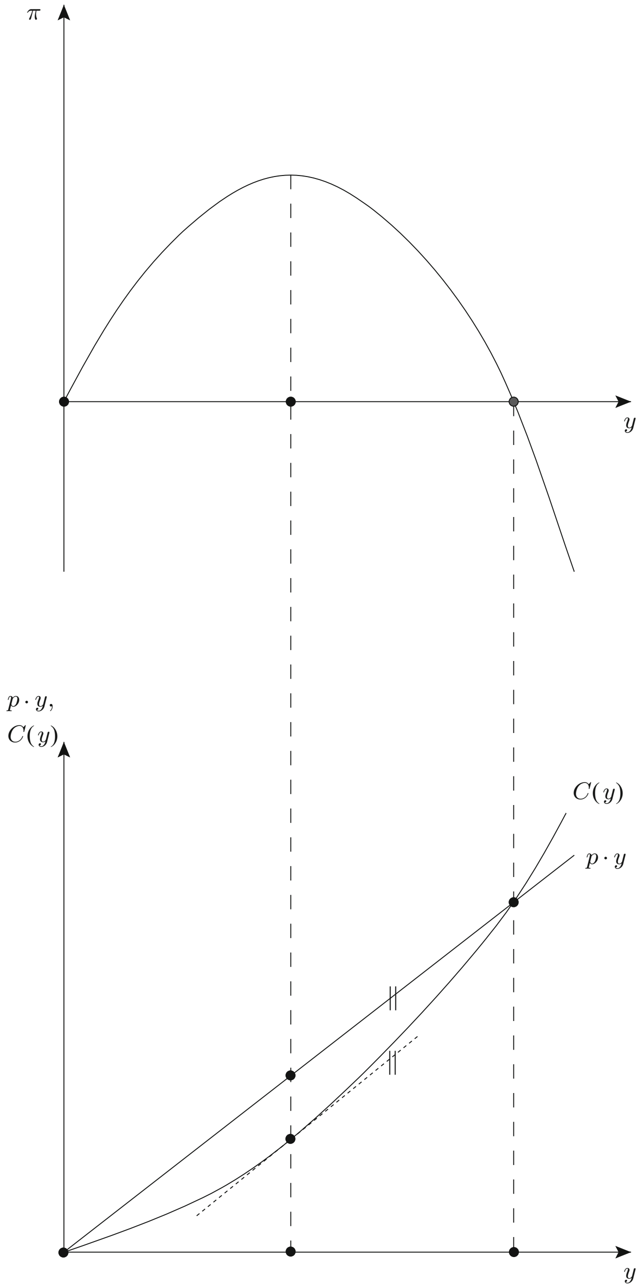

Fig. 9.1 Profits as a function of output

gets 80% of the output in 20% of the time, and the additional 20% of output in the remaining 80% of the time. This principle applies to a large number of production processes, and intellectual or physical skills are only one example. If one exploits natural resources, it is usually relatively easy at the beginning and gets increasingly difficult, when the source becomes depleted. Increasing the crop on a given piece of land is relatively easy at the beginning but, the larger the crop, the more difficult it gets to further increase it, and so on.

The above arguments used technological explanations for increasing marginal costs and this is exactly one of the reasons why I have linked costs with production functions. Given that input markets are competitive, the structure of the cost function is determined by the structure of the production function, because they are, in the one-factor example (up to a scaling factor, which is determined by input prices), inverse to each other. Increasing marginal costs exist, if the marginal product decreases, i.e. if it gets more difficult to increase production the more one is already producing.

Assume that marginal costs are decreasing. In this case, the first-order condition characterizes a minimum. What are the economic implications of this finding? Figure 9.2 illustrates this case.

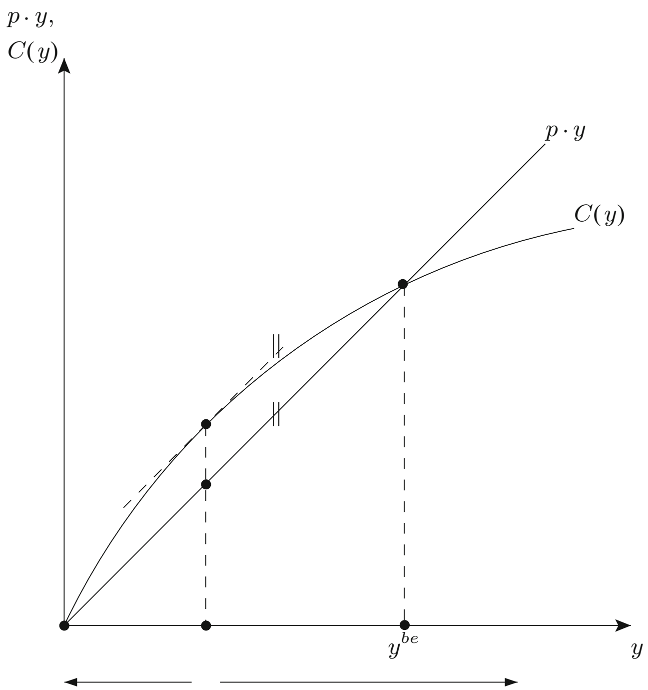

Fig. 9.2 Costs and revenues, if costs increase disproportionally with output

Figure 9.2 shows revenues (straight line) and costs (curved line) as functions of output. What one can see is that the output level, where price equals marginal costs, now characterizes the profit minimum and the marginal-cost curve is downward-sloping. At this point, the firm basically has two strategies. It can leave the market and make zero profits (arrow to the left) at $y = 0$, or it can try to grow as large as possible (arrow to the right). If the firm is successful in growing beyond the point y^{be}, it starts making profits. However, the firm should not stop here; the figure reveals that the difference between revenues and costs get larger the more the firm produces. Therefore, it is the best strategy for the firm to grow as large as possible. However, this strategy is incompatible with the assumption that the firm takes prices as given because of its smallness relative to the rest of the market. If a firm gets so large that it can serve the whole market, then it is able to influence the market price. The assumptions of perfect competition and decreasing marginal costs are logically incompatible.

The implication of this inconsistency is that perfectly competitive markets are no one-size-fits-all institution that can be used to organize economic activities. Such markets can only sustain themselves if the industry produces with the “right” type of technology, and an important number of industries do not fit into this picture.

The intermediate case, of constant marginal costs, deserves some attention, as well. Figure 9.3 shows revenues and costs as functions of output.

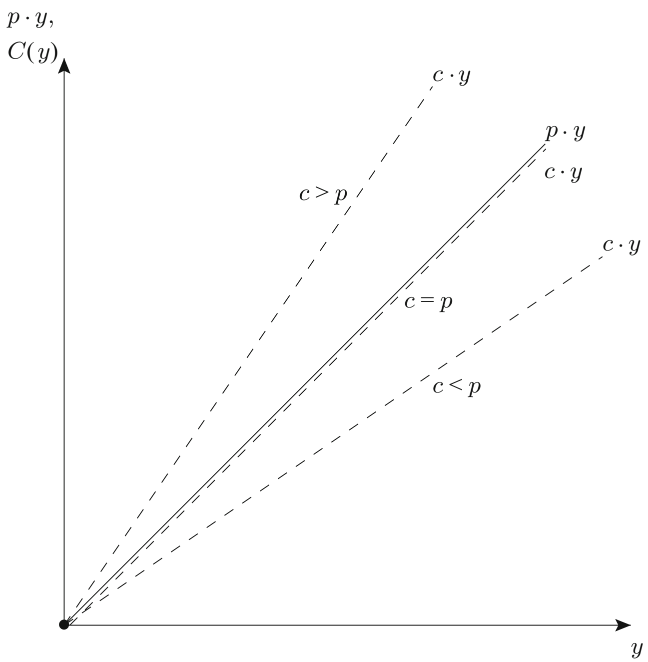

Fig. 9.3 Costs and revenues, if costs increase proportionally with output

Constant marginal costs imply that the cost function is linear. One can denote it as $C(y) = c \cdot y$ with $c > 0$. There are three possible cases: the cost function is steeper than the revenue function, $c > p$; flatter, $c < p$; or both have the same slope. The economic implications of these scenarios are straightforward: the best the firm can do, if $c > p$, is to shut down its business and to leave the market. If $c < p$, however, the opposite is the case. The firm should grow indefinitely, which is in conflict with the assumption of perfect competition. Thus, there is only one case left, where perfect competition is compatible with constant marginal costs: $c = p$. In this case, the firm is indifferent between all production levels, because it makes zero profits irrespective of how much it produces. (At some other point, I explain in detail why zero profits do not imply zero gains from trade. Zero profits imply that equity owners cannot expect a rate of return that exceeds the market interest rate for a similar investment. Hence, one can assume that the firm continues to produce, even with zero profits.)

9.3 The Short Versus the Long Run

I have argued in Chap. 8 that, depending on the time frame and the term structure of contracts, some costs of the firm can be contractually fixed and some are variable. The "price-equals-marginal-costs" rule made the implicit assumption that the firm is active on the market. However, it can always decide to leave the market and this may be a wise decision, if the losses that occur when leaving the market are smaller than the losses would be, if it stays. This statement may sound dubious at first, so one has to dig a little deeper to understand what exactly is meant by it.

Assume that a firm produces with technological fixed costs $FC > 0$ and variable costs:

$$C(y) = \begin{cases} 0, & \text{for } \; y = 0 \\ VC(y) + FC, & \text{for } \; y > 0 \end{cases},$$

and further assume that marginal costs are increasing. This situation is depicted in Fig. 9.4 with the example of a variable cost function being $VC(y) = 0.5 \cdot y^2$. Please note that this function implies that the marginal costs $MC(y) = y$ for all $y > 0$.

The horizontal line (p) is the market price and the linear monotonic line represents the marginal costs. If the firm decides to stay in the market, it will choose the quantity that equals price and marginal costs, indicated by y^* in the figure. Total revenues for an output of y^* are $p \cdot y^*$, and can be represented by the rectangular area $0pAy^*0$. Given that marginal costs are the first derivative of the variable-cost function, the triangular area $0Ay^*0$, under the marginal-cost curve, represents the variable costs. Hence, the producer surplus, $PS(y^*)$, is given by the triangular area $0pA0$ and is equal to revenues minus variable costs.

This is a general property. One may have wondered why the measure for the gains from trade of a firm is called producer surplus instead of profit. The reason is that the profit includes fixed costs, whereas producer surplus does not. One can

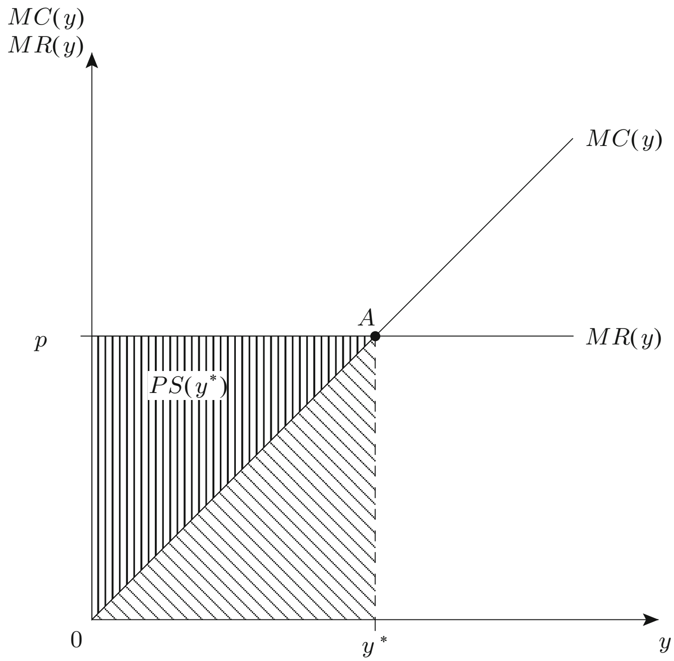

Fig. 9.4 Revenues and variable costs

establish the following relationship between profit and producer surplus:

$$\pi(y) = PS(y) - FC.$$

It follows that producer surplus and profits coincide, if fixed costs are zero, $FC = 0$. In this case, the area $0pA0$ is also the profit of the firm. However, if production requires upfront investments, then the profit is lower than the producer surplus.

The easiest way to see how technological fixed costs (remember that these costs can be avoided by producing a quantity of 0) influence profits is by adding the average-cost curve to the picture. It is equal to

$$AC(y) = \frac{VC(y)}{y} + \frac{FC}{y} = 0.5 \cdot y + \frac{FC}{y},$$

if $y > 0$. A closer look at the expression reveals that it is u-shaped: the first term is a linear, increasing function, whereas the second term is hyperbolic. Hence, the sum must be u-shaped. Is there anything else that one can say about the average-cost curve? Yes: it intersects with the marginal-cost curve at the minimum of the average costs. To understand this intuitively, think of the range over which the average-cost curve is declining. Within this range, marginal costs must be smaller than average costs: if the average is declining at a given point, then the cost of the last unit needs to be below the average costs up to this point. By the same token, if the average-cost curve is increasing, then the costs of the last unit must be higher than the average costs at any given point, because otherwise marginal costs would not have sufficient "leverage" to bring up average costs.

To see this algebraically, note that, applying the quotient rule, a necessary condition for the minimum of the average-cost curve is

$$AC'(y) = 0 \Leftrightarrow \frac{C'(y) \cdot y - C(y)}{y^2} = 0.$$

For $y > 0$, this condition can be simplified to

$$C'(y) \cdot y - C(y) = 0,$$

if one multiplies by y^2 (which is possible because $y > 0$). Dividing by y and rearranging terms gives the desired result:

$$C'(y) = \frac{C(y)}{y} \Leftrightarrow MC(y) = AC(y).$$

(The same calculation can be carried out for average variable costs.) To illustrate, Fig. 9.5 shows the marginal-cost curve and the average-cost curve for $FC = 10$.

Note that $AC(y) = C(y)/y$, or $C(y) = AC(y) \cdot y$, which means that total costs for any output y can be measured by the rectangular area $0ABy0$.

With this prerequisite, one can return to the relationship between producer surplus and profit, or the role of technological fixed costs in firm behavior. Different levels of technological fixed costs give rise to a family of average-cost curves, where higher curves correspond to higher technological fixed costs.

Figure 9.6 shows the marginal-cost curve and the family of average-cost curves for different values of FC.

There is one average-cost curve that is of particular interest: the one that intersects with the marginal-cost curve exactly where price equals marginal costs. Call this level of fixed costs FC'. This situation is represented in Fig. 9.7.

One already knows that total revenues are equal to $0pAy^*0$ at y^* and that producer surplus is equal to $0pA0$. What one knows, in addition, is that total costs are $0pAy^*0$ at y^*, so total profits are equal to zero. What is happening here is that the producer surplus is sufficient to cover the technological fixed costs. If fixed costs are lower, the firm stays in business and makes a profit. What happens, however, if the technological fixed costs are higher? In this case, the firm would end up with a negative profit or loss. Is there anything the firm can do about this loss? Yes, given that one is talking about technological fixed costs, it can *ex-ante* (before it starts the development process), anticipate that the producer surplus will be insufficient to cover the technological fixed costs, at the expected market price, and stay out of business. At this *ex-ante* stage, the total profit from staying out of business is 0, which is better than the loss that results from entering the market. This finding

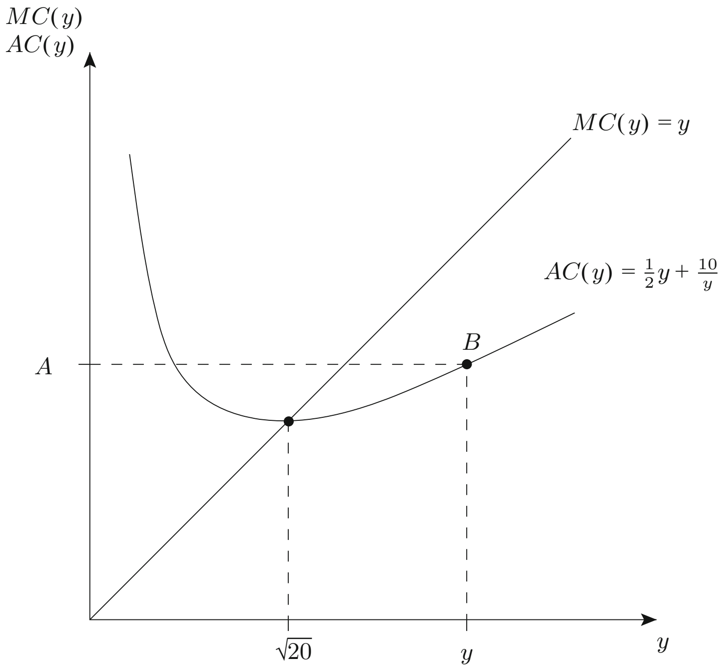

Fig. 9.5 Marginal and average costs

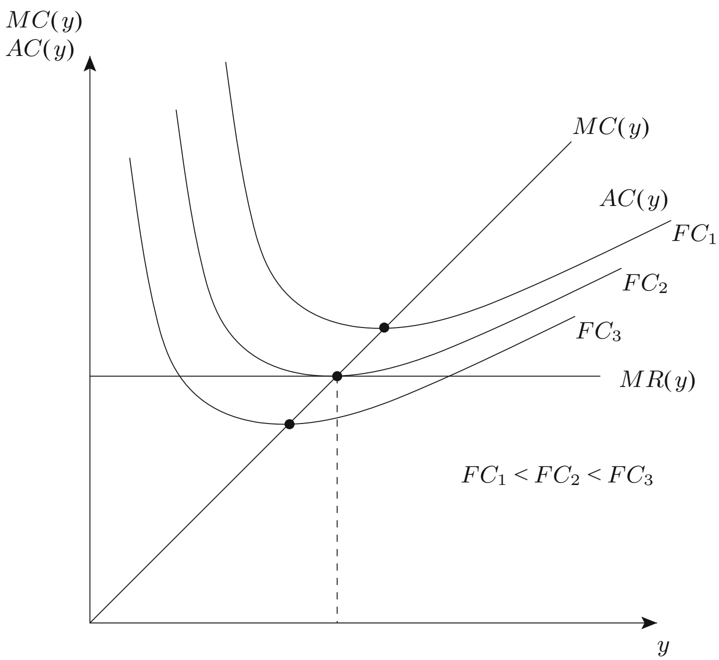

Fig. 9.6 A family of average-cost curves

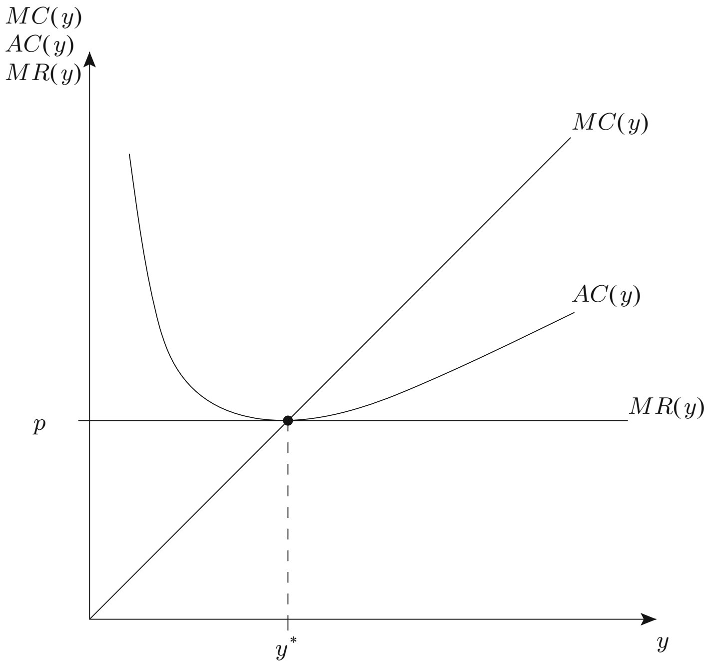

Fig. 9.7 Average costs such that profits are equal to zero

leads to an important modification of the optimal supply decision. In the long run, when all costs can be avoided by not entering the market, the optimal strategy of a competitive firm is to determine the optimal quantity, according to the price-equals-marginal-costs-rule, if the market price is (weakly) above the average costs, and to stay out of the market otherwise. The individual supply function is, therefore, identical to the inverse of the marginal-cost function, if the price is (weakly) larger than average costs.

One can now turn to a slightly modified case. Contrary to the above example, assume that the firm has already entered a contractual arrangement that turns the fixed into sunk costs, i.e. "costs" that cannot be avoided by shutting down the business. In this case, the same analysis as in Fig. 9.7 applies, but the economic consequences are different. For all levels of fixed costs $FC > FC'$ the firm makes a loss, but this loss cannot be avoided by going out of business. Hence, the best the firm can do is to minimize losses, which means sticking to the price-equals-marginal-costs rule. This rule will lead to losses in the end, but they are smaller than the loss that would occur with any other strategy, including going out of business. Figure 9.8 compares the two scenarios.

The upward-sloping function is the marginal-cost curve and the u-shaped function is the average-cost curve. The supply curve equals the section of the marginal-cost curve above the average-cost curve, if the technological fixed costs are not yet

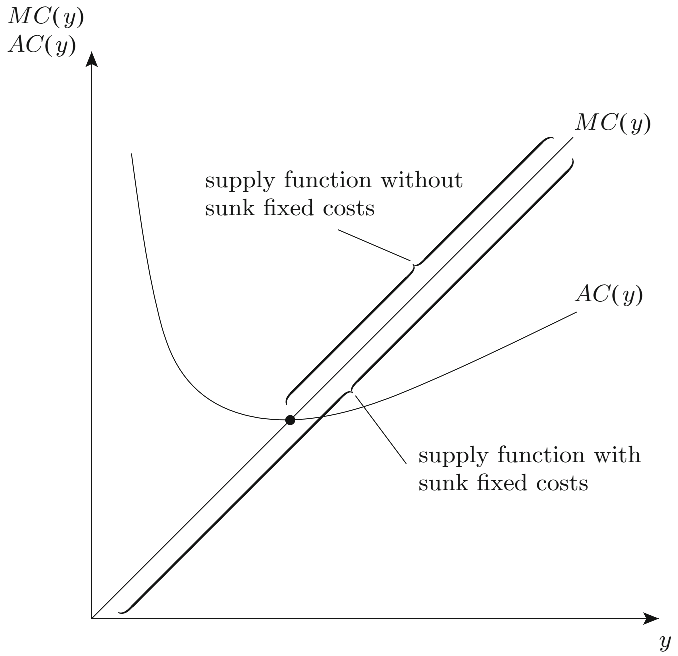

Fig. 9.8 Supply functions with and without sunk fixed costs

sunk, and it equals the complete marginal-cost function, if the technological fixed costs are sunk.

$\mathcal{L}$, $\mathcal{B}$ The above example carved out the implications of the difference between technological and contractual fixed costs. Contractual fixed costs can, however, also exist in situations where technological fixed costs are zero. Assume, for example, that a farmer has a cherry orchard with a given number of trees. The only additional input that he needs at harvest time is labor. Assume also that the quantity of cherries picked is increasing in the number of hours fruits are picked, but that the increase is declining (it gets harder and harder to pick additional cherries). At a given market wage, this "picking technology" creates increasing marginal costs of fruit picking. If the farmer can hire fruit pickers on a daily spot market, this assumption turns wage payments into variable costs. Assume, on the contrary, that a union of fruit pickers negotiated a three months dismissal protection. In this case, labor costs become fixed and sunk once the employment contract is signed. The analysis of situations like this is qualitatively identical to the analysis above and the basic understanding is simple: *costs that the firm cannot influence have no significance for the optimal behavior of the firm.* Alternatively, to put it shortly: *sunk costs are sunk.*

At this point, a remark is in order about the role that sunk costs play in standard economics. It is a generally accepted view that rational decision makers ignore sunk costs in their decisions: if one cannot influence them, they should be irrelevant for one's decisions. One calls this the *sunk-cost principle*. Generally, it is a wise and important normative principle: one should not care about the past, if one wants to make rational decisions (but one should have a look at the digression below). However, it is less clear that its predictive power in positive theory is very high. In a number of cases, people care about sunk costs, even if they should not, according to the sunk-cost principle.

An example is the empirical phenomenon of *mental accounting* that describes the tendency of individuals to keep different financial titles in different "mental accounts," and to evaluate the performance of the different titles separately, despite of the fact that a rational decision maker should aggregate them and evaluate the performance of the whole portfolio. For example, assume that someone made equal investments in two stocks. If he sold them today, stock A would have gained CHF 5,000 and stock B would have lost CHF 5,000. Assume that he has to sell a stock because he needs some extra liquidity. A rational person would take the past performance of stocks into consideration, only if he thinks that past performance is correlated with future performance, such that one can learn from the past. Otherwise, past gains and losses should be irrelevant for one's decision to sell stock A or stock B. However, empirical evidence shows that most people have a preference for selling the winning stock A, which could only be rationalized, if the good past performance is an indicator for bad future performance. Much more likely is the explanation that they hold both stocks in different mental accounts and react emotionally to realized losses and gains. Capitalizing the gains from selling stock A gives one pleasure and, at the same time, it allows one to avoid the confrontation with the pain of realizing the losses of stock B. These emotional predispositions may influence one's behavior and make it incompatible with the sunk-cost principle. The tendency to invest additional resources into losing accounts or to "throw good money after bad" is sometimes also called the *sunk-cost fallacy*.

Digression 23. Evolution, Emotions, and Sunk Costs: When Caring about Sunk Costs Can Be Beneficial Φ

It appears that deviations from the sunk-cost principle are always bad. However, if this were the case, one may wonder why humans' brains evolved in a way that make us vulnerable to the sunk-cost fallacy. Recent research in evolutionary biology challenges the theory that such behavior is necessarily bad. Take the so-called ultimatum game as an example. In this game, two players have to decide how to divide a sum of money. The first player can propose how to split the sum between the two players and the other player can then accept or reject the offer. If she accepts, the money will be split according to the proposal; if she rejects, neither player receives anything. According to the sunk-cost principle, the second player should accept any positive amount, because the proposal of the first player is in the past and cannot be influenced. Nevertheless, with this logic, a selfish player 1 should offer the minimum amount possible. Hence, the sunk-cost principle guarantees that player 2 gets almost nothing.

This prediction has been consistently tested and falsified in the laboratory. It turns out that subjects in the role of player 2 very often reject small offers, because they find them unfair or even outrageous. However, rejecting positive offers violates the sunk-cost principle. In the end one walks home without any money when one could at lease have had some. From an evolutionary point of view, however, the apparently dysfunctional emotions of anger, frustration or rage that lead one to turn down flimsy offers may play a very functional role. Within a community, reputation takes on a vital role in human interaction, because it is not unlikely for one to be in a position to do business with the same person more than once or with people who have heard about one's previous business dealings. Thus, player 2 would like to commit to a strategy that turns down bad offers because, if player 1 knows that bad offers will be turned down, he has an incentive to make better ones. The problem is, of course, how to make such an announcement credible. An important role emotions seem to be playing in regulating human interactions is exactly this: to make credible commitments possible. Assume player 2 reacts with anger and frustration to bad offers, so that he happily rejects them and player 1 knows this (either by introspection or because he knows player 2). This knowledge would motivate player 1 to make a better offer, with the consequence that the resulting allocation is more egalitarian, which gives an "emotional" player 2 a fitness advantage over a purely "rational" player 2. What this example shows is that one's behavioral dispositions and emotional reactions evolved over a long period of time and that they are usually functional adaptions to certain environments. In different environments, however, they may become dysfunctional. This is why it would be completely premature to classify the sunk-cost principle as the only rational way to make decisions; it all depends on the context.

9.4 Firm and Market Supply

This chapter has, up until this point, concentrated on the behavior of an individual firm. It has also shown that one can interpret the individual supply function as the inverse of the marginal-cost function, if the market price exceeds a certain benchmark, which is defined by the relevant average costs of the firm. It has neglected, thus far, to cover the relationship between individual and market supply. The assumption that firms seek to maximize profits allows one to say a little bit more than one already knows from Sect. 4.2.

Section 4.2 worked under the assumption that there is a given number of firms l in each market i. In this case, market supply for a good j was defined as

$$y_i(p_i, r, w) = \sum_{j=1}^{l} y_i^j(p_i, r, w).$$

Depending on market prices, fixed and sunk costs, the optimal policy of a firm can lead to positive profits, zero profits or even losses (during a period of time when contractual obligations restrain firms). Assume now that, for the given number of firms l, profits are strictly positive. Under certain conditions, such a situation is unsustainable in the long run, if firms maximize profits. Positive profits in a market attract additional firms to enter the market.

A good example is Lidl and Aldi, two German grocery stores that entered the Swiss market a few years ago. The old situation, in which two major incumbents, Coop and Migros, divided the lion's share of the Swiss market, was no longer sustainable after Switzerland signed the Bilateral Agreements II with the European Union, which became relevant for the food industry in 2004. These treaties opened the Swiss market for new entrants from the European Union and the relatively high profit margins, in fact, encouraged Aldi (in 2005) and Lidl (in 2009) to enter. $\mathcal{L}$

As the example shows, there may be legal impediments to entering a market, but there may be technological ones, as well: for example, if one has to invest in an infrastructure for the distribution of one's products, whose value depreciates if one leaves the market again. The loss in value is like a barrier to entering the market, because it defines a minimum producer surplus below which market entry is not profitable.

There may not only be barriers to enter, but also barriers to exit a market. Most of them are related to unfinished contractual obligations, which create financial liabilities even after leaving the market, as seen in the sunk-costs example above. However, there may also be technological closure costs, like shipping costs of equipment. With positive exit costs, a firm might be forced to stay in the market because the costs of leaving are higher than the operative loss.

I will focus on the extreme case, where entry and exit costs are zero, because it allows one to derive a very strong conclusion about the effects of competition. If the number of firms is fixed, profits can be positive, if the price exceeds average costs. Therefore, without market entry, it may be possible that profits are positive. Without entry and exit costs, these profits will encourage other firms to enter the market. This process will continue until they drive profits down to zero. Any other solution would be incompatible with the assumption of profit maximization. However, this is only possible if the market price is equal to the average costs of the firm, which is the situation that Fig. 9.7 illustrates.

This equilibrium is bad news for the owners of firms and good news for the general public. It is bad news for owners, because any expectation about positive profits will ultimately be discouraged, because market forces drive them down to zero. This finding illustrates Adam Smith's quote from the beginning of this chapter.

Zero profits does not mean that being in business is meaningless, as one can see from the cost equation. I will focus on two factors, capital K and labor L, for simplicity. Zero profits means that revenues $p \cdot y$ equal costs $C(K, L, r, w) = r \cdot K + w \cdot L$. Assume the owner provides all the capital and works himself. In this case, zero profits means that he cannot expect a compensation for his capital and labor that exceeds the market interest rate r and the market wage w, because all the revenues of the firm are completely used for factor payments, whose opportunity costs are evaluated at the input prices. Therefore, the owner is indifferent between investing in her own firm, renting out the capital at an interest rate r and working for his own firm or working for someone else for a wage w. Zero profit, in other words, does not mean that there are no gains from trade; it only implies that the owners of a company do not get rents larger than the current market rates.

From the point of view of the general public, zero profits are good news: they imply that production takes place at minimum average costs, because marginal costs intersect with average costs at the minimum of the average costs. As long as profits are positive or negative, the average costs of production are not at their minimum. The allocation is efficient, given the number of firms in the market, but the number of firms (or factories) is not yet optimal. Free entry and exit implies that, in the long run, even the number of firms adjusts such that goods are produced in the cheapest possible way.

$\mathcal{L}$, $\mathcal{B}$

Digression 24. The Ethics of Profit Maximization

> Profit is useful if it serves as a means towards an end that provides a sense both of how to produce it and how to make good use of it. Once profit becomes the exclusive goal, if it is produced by improper means and without the common good as its ultimate end, it risks destroying wealth and creating poverty. (Benedict XVI (2009), Caritas in Veritate)

One of the most intensely scrutinized assumptions of mainstream economics is profit maximization. Most people find it unethical, or even morally offensive, and claim that profit maximization is a major source of the problems of capitalist societies. The idea of *corporate social responsibility* (CSR) is seen as an alternative to profit maximization, which helps firms to better align their behavior with society's interests.

The debate about ethical and moral standards in business is probably as old as business itself. One of the oldest deciphered writings of significant length in the world, the Code of Hammurabi (1700s B.C.), lays down the rules of commerce and prescribes prices and tariffs, as well as penalties for noncompliance.

According to the 2001 Greenbook by the European Union, CSR is a "concept whereby companies integrate social and environmental concerns in their business operations and in their interaction with their stakeholders on a voluntary basis." In addition, since 2011 the European Union defines CSR as "the responsibility of enterprises for their impacts on society." This concept goes far beyond the narrow idea of profit maximization, which was put forward by Milton Friedman (1970): "In [a free economy] there is one and only one social responsibility of business – to use its resources and engage in activities designed to increase its profits so long as it stays within the rules of the game." This quote nicely expresses the mainstream view that normative concerns should, and can be, addressed at the level of the foundational institutions of society: the "rules of the game." One has seen examples for this approach in the preceding chapters: externalities should be internalized by the design of property rights, contract law, taxes, regulations and so on, but not by appealing to firms to voluntarily internalize them by non-profit-maximizing business practices.

Given these opposing views, is it possible to bridge them? For starters, one gets a lot of support for the so-called *Friedman doctrine* from the model of firm behavior under perfect competition, which was developed in this chapter. First, note that the existence of a complete set of competitive markets implies an ideal institutional framework which, in the language of Milton Friedman, could be understood as the perfect rules of the game. This is expressed in the First Theorem of Welfare Economics. Second, with free entry and exit, competition has the tendency to drive profits to zero. However, if profits are zero at the maximum, firms do not have much choice but to maximize them. Paying higher wages to employees or selling at lower prices simply drives firms out of business. The only exceptions to this rule are short-run profits, or a situation where entry and exit are restricted, such that profits are positive even in the long-run. However, in this case, advocates of free markets would argue that one should first try to reduce the entry- and exit barriers to the largest extent possible, in this case. Like it or not, under perfect competition, there is not much room for anything but profit maximization.

A lot of firms voluntarily choose ethically sound business practices. One has to be careful to judge these practices correctly, though. Their existence does not necessarily imply that firms incorporate other objectives than simply profits into their business models. There are a number of cases where a more comprehensive understanding of the factors that influence the adoption of these practices is necessary. For example, there are apparently cases in which ethical practices are profit maximizing in a long time horizon. Paying decent wages may motivate employees to work harder and to be loyal to the firm, thereby increasing profits. Sustainability standards may lead to higher prices, if consumers have a willingness to pay for sustainable production, and so on. In fact, a lot of proponents of CSR reduce the concept to this "enlightened self-interest" of the firm, the argument being that a lot of potential conflicts of interest between the owners and managers of firms ("shareholders") and other groups in society ("stakeholders") result from a too-narrow perspective of the shareholders. This view implicitly accepts the profit motive, but aligns it with social interests by declaring them compatible. The approach could also be called *responsible profit maximization*.

However, this is not the end of the story. One has seen that perfect competition depends on technological prerequisites, which are not always fulfilled, and that externalities may make an equilibrium inefficient. In situations like these, there is room for discussion about the adequate way to address inefficiencies and problems of sustainability, as well as distributive justice on the firm level. Here is an example: one of the major challenges of globalization is exactly the lack of a consistent global regulatory framework – the rules of the game – that create a perfectly level playing field, and institutions like the WTO or OECD are too weak to fill the holes and gaps in the playground. (Nevertheless, CSR goes beyond the problems imposed by globalization.) Nation-states even enter into race-to-the-bottom types of international competition, where they reduce taxes and social standards to attract internationally mobile capital. This type of competition can, in principle, be beneficial, if it is primarily utilized as a disciplinary device for nation states to provide public services more efficiently, but it often drives standards below the efficient level. Especially large, multinational corporations can profit from these developments and, for the foreseeable future, there is no other institutional actor able to address the ethical issues that result from these developments other than the corporations themselves. Again, like it or not, if they do not care, no one will.

References

Friedman, M. (1970). The Social Responsibility of Business is to Increase Its Profits. *The New York Times Magazine*.

Pope Benedict XVI (2009). *Caritas in Veritate*. Veritas Publications.

Smith, A. (1776)[1991]. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Everyman's Library.

Further Reading

Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic Theory*. Oxford University Press.

Varian, H. R. (1992). *Microeconomic Analysis*. Norton.