Chapter 2

The Firm

> I think business is very simple. Profit. Loss. Take the sales, subtract the costs and you get this big positive number. The math is quite straightforward. – Bill Gates, *US News and World Report*, 15 February 1993.

2.1 Basic setting

We begin with the economic problem of the firm, partly because an understanding of this subject provides a good basis for several other topics that arise later in the book, partly because the formal analysis of this problem is quite straightforward and can usually be tied into everyday experience and observation.

We will tackle the issues that arise in the microeconomic analysis of the firm in seven stages. The first four of these are as follows:

- We analyse the structure of production and introduce some basic concepts that are useful in solving the firm's optimisation problem.
- We solve the optimisation problem of the price-taking, profit-maximising firm. Along the way we look at the problem of cost-minimisation.
- The solution functions from the optimisation are used to characterise the firm's responses to market stimuli in the long and the short run.
- The analysis is extended to consider the problems confronting a multi-product firm.

The remaining three topics focus on the firm's relationship with the market and are dealt with in chapter 3.

In this chapter we will find in part a review of some standard results that you may have already encountered in introductory treatments of microeconomics, and in part introduce a framework for future analysis. I shall give a brief account

z_i	amount used of input i
q	amount of output
ϕ	production function
w_i	price of input i
p	price of output

Table 2.1: The Firm: Basic Notation

of the behaviour of a firm under very special assumptions; we then build on this by relaxing some of the assumptions and by showing how the main results carry over to other interesting issues. This follows a strategy that is used throughout the later chapters – set out the principles in simple cases and then move on to consider the way the principles need to be modified for more challenging situations and for other economic settings that lend themselves to the same type of treatment.

2.1.1 The firm: basic ingredients

Let us introduce the three main components of the problem, the technology, the environment, and economic motivation.

Technology

You may well be familiar with the idea of a production function. Perhaps the form you have seen it before is as a simple one-output, two-input equation: $q = F(K, L)$ ("quantity of output = a function of capital and labour"), which is a convenient way of picking up some of the features that are essential to analysing the behaviour of the firm.

However, we shall express the technological possibilities for a firm in terms of a fundamental *in*equality specifying the relationship between a single output and a vector of m inputs:

$$q \leq \phi(\mathbf{z}) \tag{2.1}$$

Expression (2.1) allows for a generalisation of the idea of the production relation. Essentially the function ϕ tells us the maximum amount of output q that can be obtained from the list of inputs $\mathbf{z} := (z_1, z_2, .., z_m)$; putting the specification of technological possibilities given in the form (2.1) allows us to:

- handle multiple inputs,
- consider the possibility of inefficient production.

On the second point note that if the "=" part of (2.1) holds we shall call production *technically efficient* – you cannot get any more output for the given list of inputs $\mathbf{z}$.

The particular properties of the function ϕ incorporate our assumptions about the "facts of life" concerning the production technology of the firm.

Working with the single-product firm makes description of the "direction of production" easy. However, sometimes we have to represent multiple outputs, where this specification will not do – see section 2.5 below where we go further still in generalising the concept of the production function.

Environment

We assume that the firm operates in a market in which there is pure competition. The meaning of this in the present context is simply that the firm takes as given a price p for its output and a list of prices $\mathbf{w} := (w_1, w_2, .., w_m)$ for each of the m inputs respectively (mnemonic – think of w_i as the "wage" of input i).

Of course it may be interesting to consider forms of economic organisation other than the market, and it may also be reasonable to introduce other constraints in addition to those imposed by a simple specification of market conditions – for example the problem of "short-run" optimisation, or of rationing. However, the standard competitive, price-taking model provides a solid analytical basis for a careful discussion of these other possibilities for the firm and for situations where a firm has some control over the price of output p or of some of the input prices w_i.

Motivation

Almost without exception we shall assume that the objective of the firm is to *maximise profits*: this assumes either that the firm is run by owner-managers or that the firm correctly interprets shareholders' interests.[1]

Within the context of our simplified model we can write down profits in schematic terms as follows:

firm's profits	=	sales revenue	–	purchases of inputs

More formally, we define the expression for profits as

$$\Pi := pq - \sum_{i=1}^{n} w_i z_i \tag{2.2}$$

Before we go any further let us note that it seems reasonable to assume that ϕ in (2.1) has the property:

$$\phi(\mathbf{0}) = 0 \tag{2.3}$$

which in plain language means both that the firm cannot make something for nothing and that it can always decide to shut up shop, use no inputs, produce no output, and thus make zero profits. Therefore we do not need to concern ourselves with the possibility of firms making negative profits (tactful name for losses) in the profit-maximisation problem.[2]

[1] What alternative to profit-maximisation might it be reasonable to consider?

[2] In real life we come across firms reporting losses. In what ways would our simplified model need to be extended in order to account for this phenomenon?

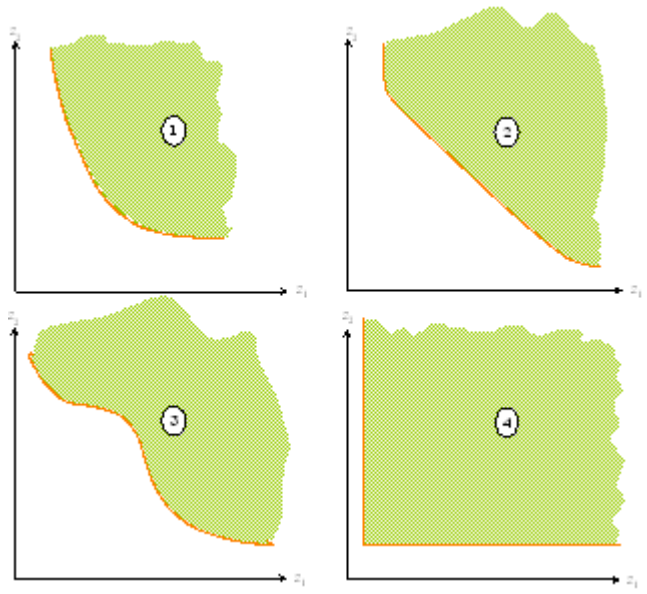

Figure 2.1: Input requirement sets for four different technologies

2.1.2 Properties of the production function

Let us examine more closely the production function given in (2.1) above. We will call a particular vector of inputs a *technique*. It is useful to introduce two concepts relating to the techniques available for a particular output level q:

1. Pick some arbitrary level of output q: then the *input-requirement set* for the specified value q is the following set of techniques:

$$Z(q) := \{\mathbf{z} : \phi(\mathbf{z}) \geq q\}. \tag{2.4}$$

2. The *q-isoquant* of the production function ϕ is the contour of ϕ in the space of inputs

$$\{\mathbf{z} : \phi(\mathbf{z}) = q\}. \tag{2.5}$$

Clearly the q-isoquant is just the boundary of $Z(q)$. Although you may be familiar with the isoquant and the input requirement-set Z may seem to be a novelty, the set Z is, in fact, useful for characterising the fundamental properties of the production function and the consequences for the behaviour of the optimising firm. Certain features of shape of Z will dictate the general way in which the firm responds to market signals as we will see in section 2.3 below.

In a 2-input version of the model Figure 2.1 illustrates four possible shapes of $Z(q)$ corresponding to different assumptions about the production function. Note the following:

- An isoquant can touch the axis if one input is not essential.
- An isoquant may have flat segments (case 2 in Figure 2.1). We can interpret this as locally perfect substitutes in production.
- The convexity of $Z(q)$ implies that production processes are, in some sense, divisible. To see this, do the following with cases 1, 2 or 4 in Figure 2.1: take any two vectors $\mathbf{z}'$ and $\mathbf{z}''$ that lie in $Z(q)$; draw the straight line between them; any point on this line clearly also belongs to $Z(q)$ and such a point can be expressed as $t\mathbf{z}' + [1-t]\mathbf{z}''$ where $0 < t < 1$; what you have established is that if the production techniques $\mathbf{z}'$ and $\mathbf{z}''$ are feasible for q, then so too is a mixture of them (half one and half the other, say).[3] However, this does not work everywhere in case 3 (check the part of Z where there is a "dent"). Here a mixture of two feasible techniques may lie outside Z: nonconvexity implies that there is some indivisibility in the production process.
- An isoquant may have "kinks" or corners: (case four).

Marginal Rate of Technical Substitution

Where ϕ is differentiable (i.e. at points on the isoquant other than kinks) we shall often find it convenient to work with the slope of the isoquant, which is formally defined as follows:

Definition 2.1 *The* marginal rate of technical substitution *of input i for input j is given by*

$$\text{MRTS}_{ij} := \frac{\phi_j(\mathbf{z})}{\phi_i(\mathbf{z})}$$

In this definition and elsewhere we use subscripts as a shorthand for the appropriate partial derivative. In this case $\phi_i(\mathbf{z})$ means $\partial\phi(\mathbf{z})/\partial z_i$.

The MRTS reflects the "relative value" of one input in terms of another from the firm's point of view. The particular value of the MRTS for inputs $\left(z_1^0, z_2^0\right)$ is represented in Figure 2.2 by the slope at point $\mathbf{z}^\circ$; the slope of the ray through $\mathbf{z}^\circ$ represents the corresponding *input ratio* z_2/z_1 at this point.

Elasticity of substitution

We can use this idea to characterise the shape of the isoquant. Consider the question: how responsive is the firm's production technology to a change in this relative valuation? This may be made precise by using the following definition.

[3] A firm has offices in London and New York. Fractional units of labour can be employed in each place (part-timers can be hired) and the headquarters could be in either city. The minimum viable office staff is 1 full-time employee and the minimum size of headquarters is 3 full-timers. Sketch the isoquants in this case and explain why $Z(q)$ is not convex.

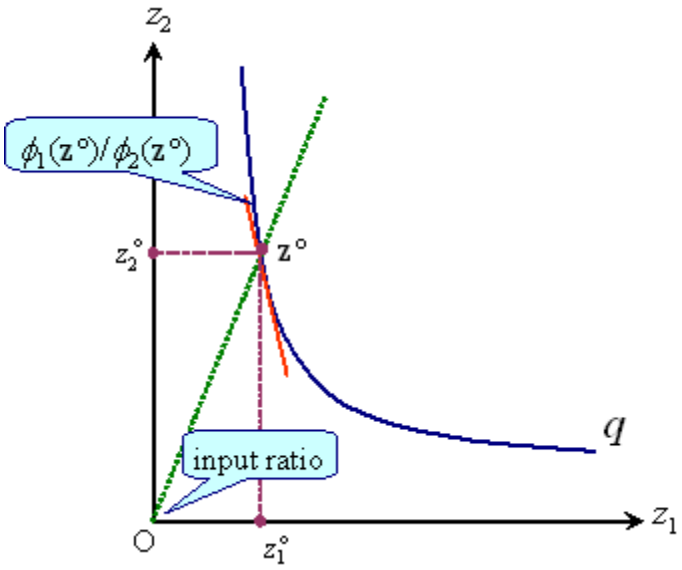

Figure 2.2: Marginal rate of technical substitution

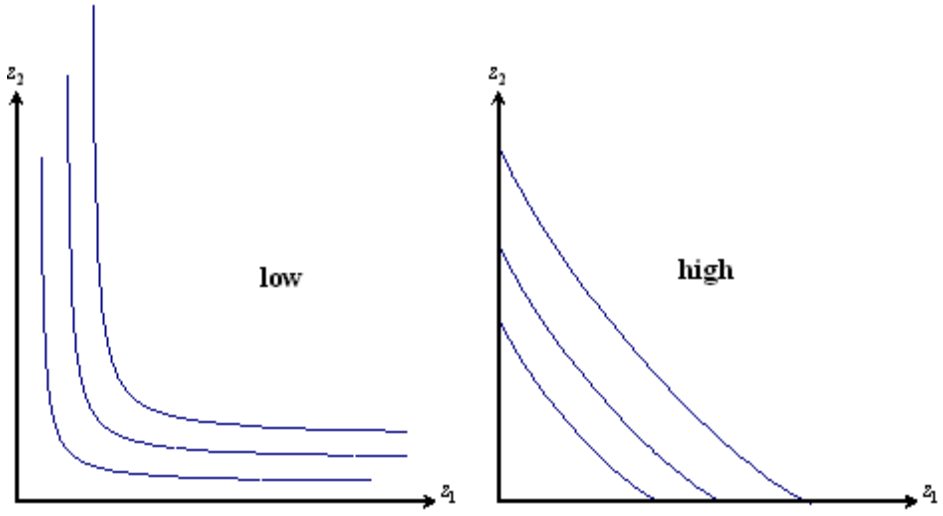

Figure 2.3: Low and high elasticity of substitution

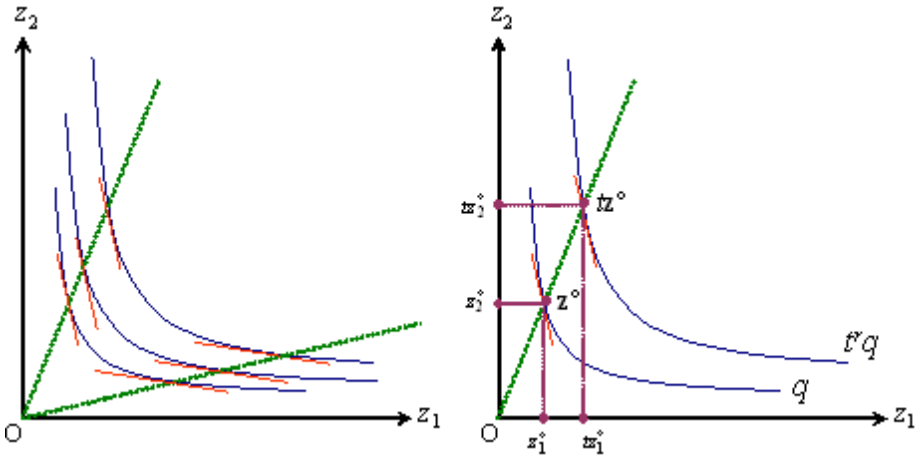

Figure 2.4: Homothetic and homogeneous functions

Definition 2.2 *The* elasticity of substitution *is*

$$\sigma_{ij} := -\frac{\partial \log (z_j/z_i)}{\partial \log \left(\phi_j(\mathbf{z})/\phi_i(\mathbf{z})\right)}. \tag{2.6}$$

Notice that $\sigma_{ij} \geq 0$ and that (2.6) has the simple interpretation

$$\frac{\text{proportional change in input ratio}}{\text{proportional change in MRTS}}$$

(in absolute terms).[4] Higher values of σ mean that the production function is more "flexible" in that there is a proportionately larger change in the production technique in response to a given proportionate change in the implicit relative valuation of the factors: Figure 2.3 illustrates isoquant maps for two cases, where σ is low (large changes in the MRTS are associated with small changes in the input ratio) and where σ is high (small changes in the MRTS are associated with large changes in the input ratio).

We can build up an entire family of isoquants corresponding to all the possible values of q and there may be a wide variety of potentially interesting forms that the resulting map might take.

Homothetic and homogenous production functions

For many purposes it is worth considering further restrictions on the function ϕ that have convenient interpretations. The left-hand half of Figure 2.4 illustrates the case of *homothetic* contours: each isoquant appears like a photocopied

[4] Show that $\sigma_{ij} = \sigma_{ji}$. You may find the material on page 496 useful.

enlargement; along any ray through the origin all the tangents have the same slope so that the MRTS depends only on the relative proportions of the inputs used in the production process. The right-hand half of Figure 2.4 illustrates an important subcase of this family – *homogeneous* production functions – for which the map looks the same but where the labelling of the contours has to satisfy the following rule: for any scalar $t > 0$ and any input vector $\mathbf{z} \geq 0$:

$$\phi(t\mathbf{z}) = t^r \phi(\mathbf{z}), \tag{2.7}$$

where r is a positive scalar. If $\phi(\cdot)$ satisfies the property in (2.7) then it is said to be homogeneous of degree r. Clearly the parameter r carries important information about the way output responds to a proportionate change in all inputs together: if $r > 1$, for example then doubling more inputs will more than double output.

Returns to scale

However, homogenous functions, although very convenient analytically, are obviously rather special. It is helpful to be able to classify the effect of changing the scale of production more generally. This is done using the following definition:

Definition 2.3 *The production function ϕ exhibits*

1. increasing returns to scale (IRTS) *if, for any scalar $t > 1$:*

$$\phi(t\mathbf{z}) > t\phi(\mathbf{z}) \tag{2.8}$$

2. decreasing returns to scale (DRTS) *if, for any scalar $t > 1$:*

$$\phi(t\mathbf{z}) < t\phi(\mathbf{z}) \tag{2.9}$$

3. constant returns to scale (CRTS) *if, for any positive scalar t:*

$$\phi(t\mathbf{z}) = t\phi(\mathbf{z}) \tag{2.10}$$

Figures 2.5 to 2.7 illustrate production functions with two inputs and a single output corresponding to each of these three cases. In each case the set of points on or "underneath" the tent-like shape represent feasible input-output combinations. Take a point on the surface such as the one marked in each of the three figures:

- Its vertical coordinate gives the maximum amount of output that can be produced from the input quantities represented by its (z_1, z_2) coordinates.
- The dotted path through this point in each figure is the *expansion path*; this gives the output and input combinations as (z_1, z_2) are varied in the same proportion (for example variations along the ray through the origin

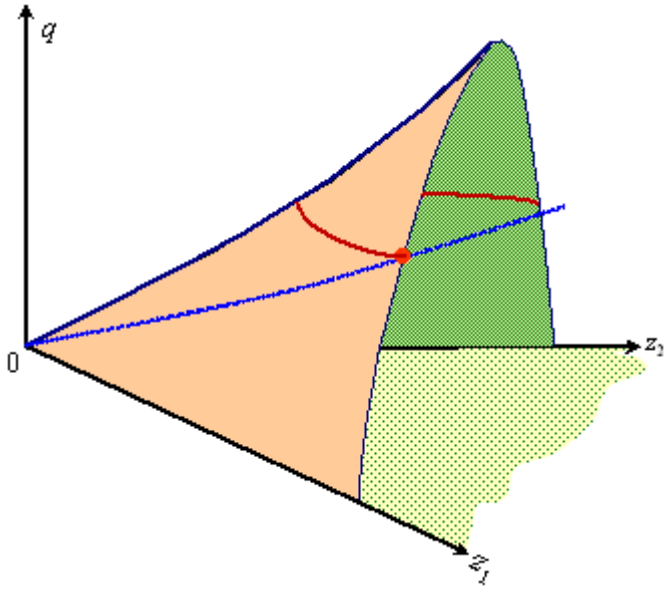

Figure 2.5: An IRTS production function

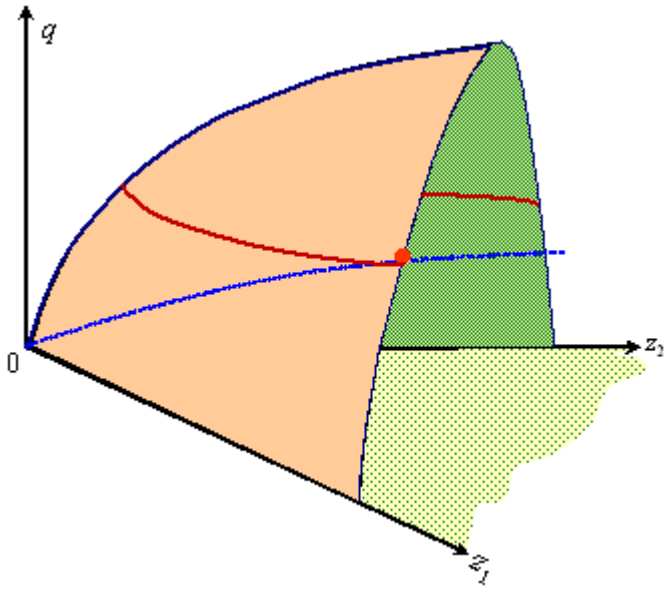

Figure 2.6: A DRTS production function

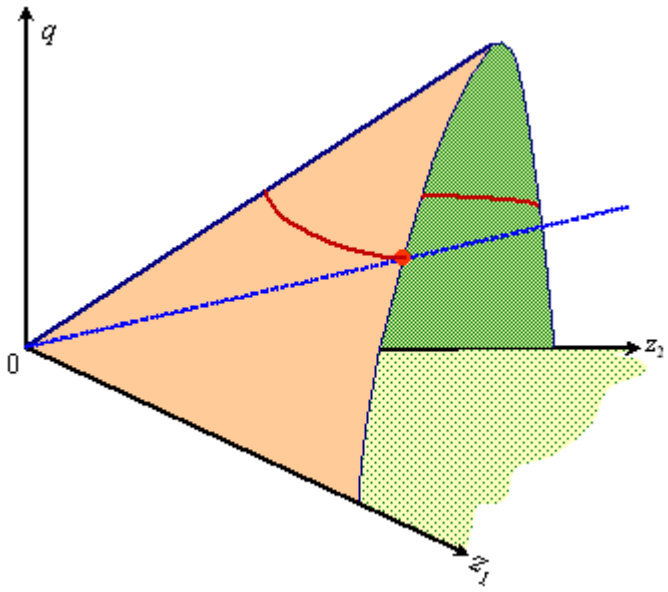

Figure 2.7: A CRTS production function

in the 2-dimensional Figure 2.2).[5] In the simple constant returns to scale production function the expansion path is itself a ray through the origin (Figure 2.7); in the IRTS and DRTS cases this path is clearly curved.

- The solid curve through this point in each figure is a *contour* of ϕ; project this contour down into the (z_1, z_2)-plane (the "floor" of the diagram) and you get the isoquant.

Of course one could specify localised increasing returns to scale by limiting the range of values of t for which (2.8) is true – likewise for decreasing or constant returns to scale; quite a common assumption is that for small-scale production (low values of z_1 and z_2) IRTS is true while for large scale operations DRTS is true. Furthermore it is easy to check that if ϕ is a concave function all the sets $Z(q)$ are convex and returns to scale are constant or decreasing everywhere.

Marginal product

Now consider the relationship between output and one input (z_1 let us say) whilst all the other inputs are kept at some fixed level. We could do this in Figure 2.7, for example, by picking an arbitrary z_2 value and then slicing through the tent-shape in a plane parallel to $q0z_1$. This would give a shape

[5] In the special case of homogeneous production functions what are the values of r that correspond to increasing/constant/decreasing returns to scale?

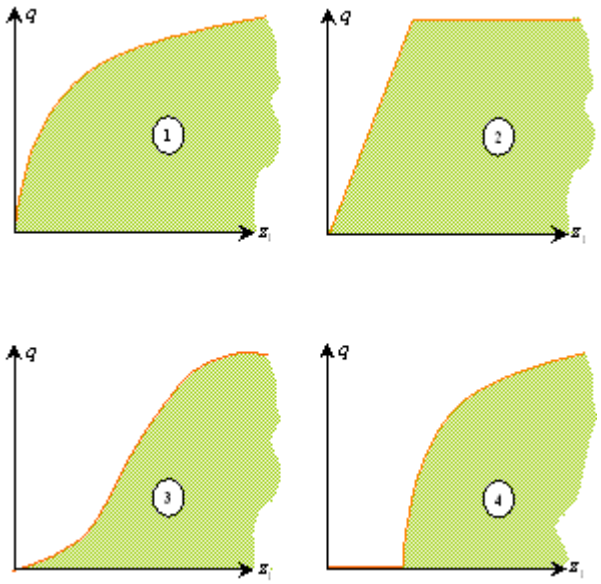

Figure 2.8: Four different technologies

such as the Case 1 in Figure 2.8.[6] Cases 2-4 in Figure 2.8 illustrate the same type of diagram for three other production functions.[7] We can use this view of the production function to depict another very useful concept, shown in Figure 2.9.

Definition 2.4 *The* marginal product *of input i is the derivative (where it is defined) of the production function.*

Of course the concept of marginal product was already implicit in the Definition 2.1 earlier: it represents the "value" to the firm of an input – measured in terms of output.

2.2 The optimisation problem

We could now set out the firm's objectives in the form of a standard constrained optimisation problem. To do this we would specify a Lagrangean incorporating profits (2.2), and the production constraint (2.1). However it is more illuminating to adopt a two-stage approach to solving the firm's optimisation problem:

[6] Sketch 3-D diagrams like the one above that will correspond to Cases 2 to 4 in Figure 2.8.

[7] Assume constant returns to scale: then two of the four cases in Figure 2.1 correpond to two of the four cases in Figure 2.8. Which are they? Suggest a simple formula for each of the two production functions that would yield these forms.

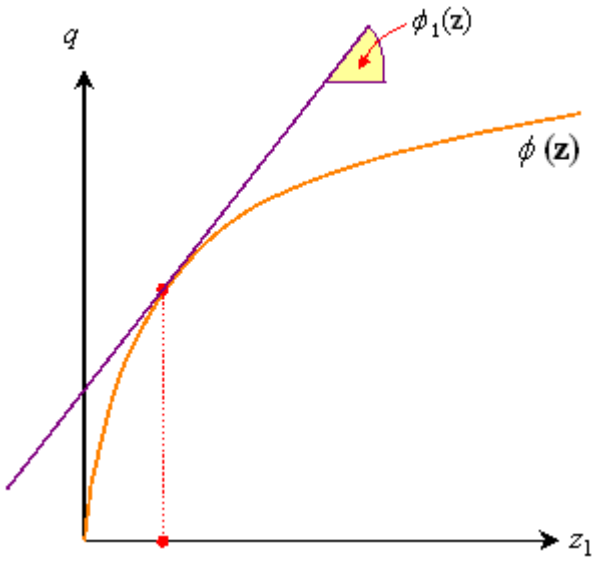

Figure 2.9: The marginal product

1. *Cost Minimisation.* For any specified output level q, find the combination of inputs that will minimise the cost of producing q for known input prices $\mathbf{w}$.

2. *Output Optimisation.* Once the appropriate input policy conditional upon an arbitrary output level is known, choose the appropriate output level.

In stage 1 we notionally *fix* the output level at some arbitrary level q as in Figure 2.1; in stage 2 the output level becomes endogenous. Why go via this roundabout route? There are two reasons. First, it neatly compartmentalises two aspects of the firm's activities that have an intuitive independent rationale; for example the stage-2 problem is a self-contained topic often presented in introductory texts. Second, the stage-1 problem is highly "portable:" we will see later examples of this approach to the solution of microeconomic problems that are in effect just a simple translation of the firm's cost-minimisation problem.

2.2.1 Optimisation stage 1: cost minimisation

The essence of the problem can be set out simply in terms of just two inputs: we can represent it diagrammatically as in Figure 2.10. Two important points to note about this diagram:

- Consider a line drawn with slope w_1/w_2 in this diagram. By definition

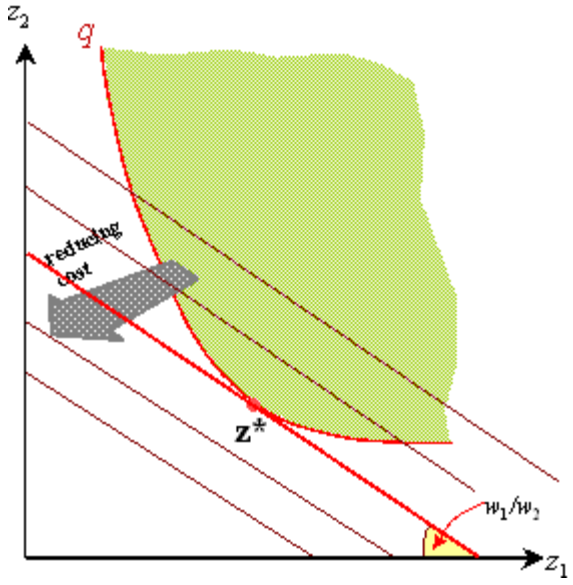

Figure 2.10: Cost minimisation

this has the equation:

$$w_1 z_2 + w_2 z_2 = \textbf{constant} \tag{2.11}$$

In other words, all the points lying on such a line represent input combinations that require the same financial outlay by the firm. For this reason such a line is known as an *isocost* line.

- Shift an isocost line up and cost goes up: you just change the constant on the right-hand side of (2.11).

Intuitively the cost-minimisation problem for a given output q involves reaching the lowest isocost line subject to staying within the input-requirement set $Z(q)$. Formally we can represent the cost-minimisation problem as that of minimising the Lagrangean:

$$\mathcal{L}(\mathbf{z}, \lambda; \mathbf{w}, q) := \sum_{i=1}^{m} w_i z_i + \lambda \left[q - \phi(\mathbf{z}) \right] \tag{2.12}$$

for some specified output level q, and for given input prices $\mathbf{w}$, subject to the restrictions that $z_i \geq 0$ for every input i, where λ is the Lagrange multiplier associated with the constraint (2.1).

Differentiating (2.12) with respect to z_i we can derive the first-order conditions (FOC) for a minimum . Let $\mathbf{z}^*$ denote the vector of cost-minimising

inputs that emerges in the solution to (2.12); if input i is used in strictly positive amounts at the optimum then the FOC implies:

$$\lambda^* \phi_i(\mathbf{z}^*) = w_i \tag{2.13}$$

More generally we have:

$$\lambda^* \phi_i(\mathbf{z}^*) \leq w_i \tag{2.14}$$

for every i where the "$<$" part applies only if $z_i^* = 0$. Likewise, differentiating (2.12) with respect to λ, we would find

$$q = \phi(\mathbf{z}^*). \tag{2.15}$$

The general condition for a maximum is actually

$$q \leq \phi(\mathbf{z}^*). \tag{2.16}$$

where the "$<$" part applies only if $\lambda^* = 0$. However, conditions (2.13, 2.14) imply that the Lagrange multiplier λ^* must be positive at the optimum,[8] and so we actually do have (2.15) – production must be technically efficient.[9] From all of this we can deduce that if cost-minimisation requires a positive amount of input i then for any other input j:[10]

$$\frac{\phi_j(\mathbf{z}^*)}{\phi_i(\mathbf{z}^*)} \leq \frac{w_j}{w_i} \tag{2.17}$$

with equality in (2.17) if input j is also used in positive amounts. So in the case where cost-minimising amounts of both inputs are positive we have:

$$\boxed{\text{MRTS}} = \boxed{\begin{array}{c}\text{input}\\ \text{price}\\ \text{ratio}\end{array}}$$

Drawing all these remarks together we have established the following result:

Theorem 2.1 (Proporties of the minimum-cost solution) *(a) The cost-minimising output under perfect competition is technically efficient. (b) For any two inputs, i, j purchased in positive amounts $MRTS_{ij}$ must equal the input price ratio w_j/w_i. (c) If i is an input that is purchased, and j is an input that is not purchased then $MRTS_{ij}$ will be less than or equal to the input price ratio w_j/w_i.*

[8] Explain why this implies that λ^* must be positive in non-trivial cases.

[9] Provide an intuitive argument to show (2.15). Hint: Suppose that at $\mathbf{z}^*$ the strict *inequality* part of (2.1) were true; show that you could then find a feasible input vector that is cheaper for the firm.

[10] (a) Draw a figure illustrating the corner solution in (b) Interpret this first-order condition using the concept of the firm's "relative value" of one input in terms of another from the firm's point of view (see page 13) (i) in the case where "$<$" holds in (2.17), (ii) in the case where "$=$" holds in (2.17)

As the earlier discussion implies, the solution may be at a corner, and it may not be unique: this all depends on the shape of the input-requirement set $Z(q)$, as we will see later.

We can express the inputs that satisfy (2.15) and (2.17) in terms of the specified output level q and the input-price vector $\mathbf{w}$. We shall write this solution as follows:

$$z_i^* = H^i(\mathbf{w}, q) \tag{2.18}$$

for inputs $i = 1, ..., m$. Think of the relationship H^i as *the conditional demand* for input i – demand that is conditional upon the level q. We shall discuss a number of aspects of this relationship – in particular the conditions under which H^i is a genuine single-valued function – after we have considered some other important features of the optimum (but you have to wait until chapter 4 on the consumer to see why the letter H is used...).

2.2.2 The cost function

We can also write the minimised cost that is the solution to (2.12) as a function of q and $\mathbf{w}$. This will prove to be a valuable concept that has applications not only throughout the rest of our discussion of the theory of the firm, but also in other areas of economic theory, such as consumer optimisation.

Definition 2.5 *The firm's* cost function *is a real-valued function C of input prices and the output level such that:*

$$C(\mathbf{w}, q) := \min_{\{\mathbf{z} \geq 0, \phi(\mathbf{z}) \geq q\}} \sum_{i=1}^{m} w_i z_i \tag{2.19}$$

$$= \sum_{i=1}^{m} w_i H^i(\mathbf{w}, q) \tag{2.20}$$

The meaning of the cost function is as follows. Given a specified value for the price of each input and for the level of output, what is the minimum outlay that the firm requires in order to purchase the inputs? Because the function C is derived from a process of cost minimisation, it possesses a number of very useful properties.

First, C must be *strictly increasing* in at least one of the input prices and, if the production function is continuous, C must be strictly increasing in output too: if this were not so then you could either use less of all inputs to get the same level of output, or get more output for the same expenditure on inputs; either way, you clearly would not be at a cost-minimising point. For much the same sort of reason we can see that C cannot be decreasing in any of the w_i.[11]

Second, we can see from (2.17) that a 10 percent increase in both input prices w_1 and w_2 would not change the optimal input levels z_i^* and z_2^*; so by how much would the minimised cost, $w_1 z_1^* + w_2 z_2^*$ have increased? Obviously

[11] C could be constant in some of the w_i. Why?

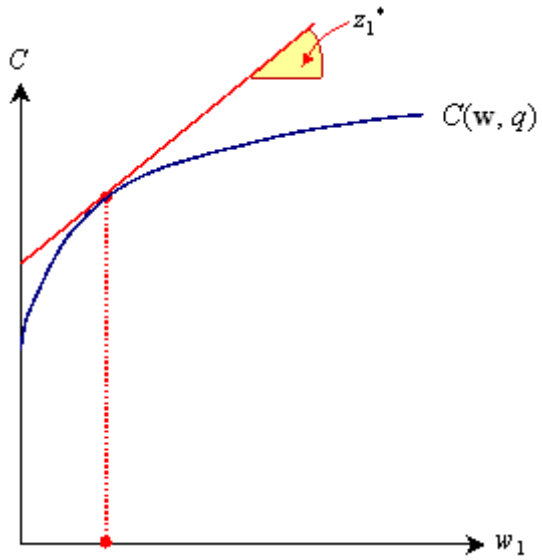

Figure 2.11: Cost and input price

10 percent. The argument easily extends to m inputs and an arbitrary rescaling of all input prices.

Third, the cost function must be *concave in prices*, as illustrated by the one-input snapshot that is illustrated in Figure 2.11: note that this a general result and does not depend on any special properties of the production function ϕ.[12]

Fourth, imagine that you are employing a thousand hours of labour at the cost-minimising point: by how much would your firm's costs increase if there was in infinitesimal increase in the wage paid to labour (say one penny an hour)? By how much would your costs have gone up had you been employing 1200 hundred units of labour at the cost-minimising point? 1200 pence? If your intuition is sharp you should have spotted that the rate of increase of cost with respect to input price equals the amount of units of that input that you employ at the optimum – a property of the cost function that is known as *Shephard's Lemma*.[13]

All these features can be summarised as follows (a proof is provided in Appendix C):

Theorem 2.2 (Properties of the cost function) *The competitive firm's cost function C in (2.19) is nondecreasing and continuous in $\mathbf{w}$, homogeneous of degree one in $\mathbf{w}$ and concave in $\mathbf{w}$. It is strictly increasing in at least one w_i.*

[12] Show that the cost function must be concave using Remark A.4 in Appendix A.

[13] Prove this in the special case where $\mathbf{z}^*$ is unique and strictly positive (Hint: differentiate (2.20) with respect to w_i and use the first-order conditions).

If the production function is continuous then C is strictly increasing in q. At every point where the differential is defined

$$\frac{\partial C(\mathbf{w}, q)}{\partial w_i} = z_i^* \tag{2.21}$$

the optimal demand for input i.

For a couple of further points of interest we introduce the concepts of *average cost* $C(\mathbf{w}, q)/q$, and of *marginal cost* $C_q(\mathbf{w}, q)$. There is a neat and very useful relationship between the "returns-to-scale" property of the production function ϕ and the behaviour of average cost: decreasing returns to scale imply rising average cost[14] and *vice versa*; constant returns to scale imply constant average cost. Also rising average cost implies that marginal cost is above average cost; falling average cost implies that marginal cost is below average cost.[15] Furthermore, consider the impact of an increase in the specified level of output on the cost minimisation problem. Noting that (2.15) holds at the optimum, we must have

$$C(\mathbf{w},q) = \sum_{i=1}^{m} w_i z_i^* + \lambda^* \left[q - \phi(\mathbf{z}^*)\right]. \tag{2.22}$$

Equation (2.22) leads to the following very useful general result on marginal cost (see Appendix C):[16]

$$C_q(\mathbf{w}, q) = \lambda^* \tag{2.23}$$

To see why we get this result, put the question: "how much would the firm be prepared to pay for an infinitesimal relaxation of the output target in (2.12) from to q to $q - \triangle q$?" The intuitive answer to this is: "an amount that is just equal to the extra cost of producing $\triangle q$." In other words, in the neighbourhood of the optimum, the appropriate "value" of the constraint in (2.12) – the Lagrange multiplier – is the marginal cost of output at q

2.2.3 Optimisation stage 2: choosing output

Using the cost function we can now set out the problem of finding optimal output. What we do is simply substitute $C(\mathbf{w}, q)$ back into (2.2). Then the problem becomes:

$$\max_{\{q \geq 0\}} \; pq - C(\mathbf{w}, q) \tag{2.24}$$

[14] Prove this. Hint: draw a pair of isoquants at $\overline{q}$ and $t\overline{q}$; for a given input-price ratio mark in the cost-minimising input combination on the $t\overline{q}$-isoquant and draw a ray through this point; find the point where this ray intersects $\overline{q}$-isoquant and work out the input bill at this point; then use the definition of the cost function.

[15] Show this.

[16] Show this in the special case where $\mathbf{z}^*$ is unique and strictly positive (Hint: differentiate (2.12) or (2.20) with respect to q and use the first-order conditions. Also check the results on page 515).

The first-order condition for this maximisation problem yields an optimum quantity q^* where

$$\left.\begin{array}{l} p = C_q(\mathbf{w}, q^*) \text{ if } q^* > 0, \\ p \leq C_q(\mathbf{w}, q^*) \text{ if } q^* = 0. \end{array}\right\} \tag{2.25}$$

In other words product price is less than or equal to marginal cost at the optimum.

A necessary condition for a maximum of (2.24) is that its second derivative with respect to q should be negative or zero in the neighbourhood of q^*. Working this out we find that this implies:

$$C_{qq}(\mathbf{w}, q) \geq 0 \tag{2.26}$$

So the optimum must be on a constant or rising portion of the marginal cost curve. However we must also take into account the obvious restriction that no firm will stay in business if it makes a loss.[17] Clearly this requires

$$pq - C(\mathbf{w}, q) \geq 0, \tag{2.27}$$

which we may rewrite as

$$\frac{C(\mathbf{w}, q)}{q} \leq p, \tag{2.28}$$

which in plain language says that average cost must not exceed product price at the optimum.

Once again we can, in principle, express the optimal supply of output as a function of the exogenously given variables in the problem by solving for q^* from the first-order condition (2.25); let us think for a moment about this supply relationship. Suppose that there is some value of output $\underline{q}$ at which marginal cost equals average cost. If marginal cost is strictly greater than average cost (to the right of $\underline{q}$),[18] and if marginal cost is rising then there is a one-to-one relationship between price p and optimal output; if marginal cost is less than average cost (to the left of $\underline{q}$), then the firm will produce no output; if marginal cost equals average cost then the firm is indifferent between producing $\underline{q}$ and producing nothing at all – see Figure 2.12. So there may be more than one profit-maximising output level for a single value of $p = \underline{p}$. We shall develop this point later, but for the moment, let us set it aside and return to the overall optimisation problem of the firm.

2.2.4 Assembling the solution

Let us now see what we get when we put together the solutions to the two component problems, cost-minimisation and output optimisation. The main result is as follows:

[17] We have ruled out $\Pi < 0$, but what would be likely to happen in a market if $\Pi > 0$? See page 57.

[18] What must be true about the production function ϕ for such a q to exist?

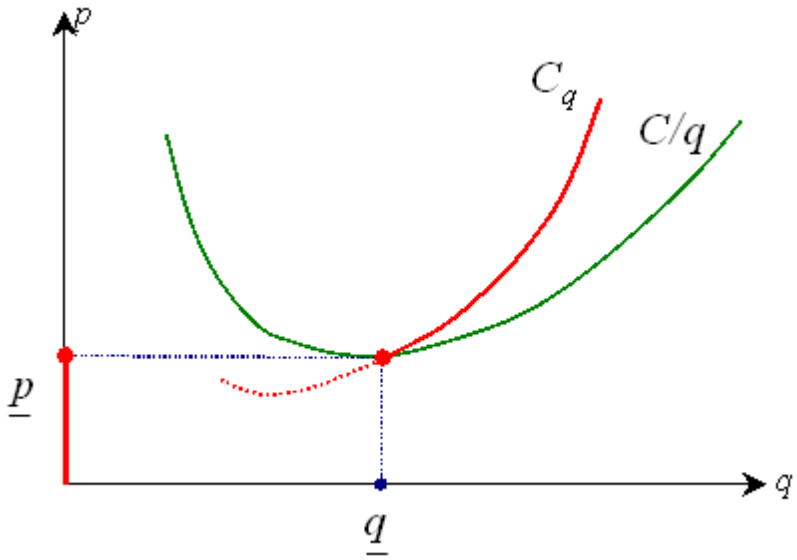

Figure 2.12: Optimal output may be multivalued

Theorem 2.3 (Marginal products and input prices) *At the profit-maximising technique, for any input the value of the marginal product of the input must be no greater than the price of that input. If the input is purchased in positive amounts, the value of its marginal product must equal its price.*

The proof of this result requires no more than gathering together some points that we already know: from expression (2.13) in the cost-minimisation problem we know that λ times the marginal product of i must be less than or equal to w_i; our discussion of the cost function revealed that λ must be marginal cost; from the optimisation of output problem we know that marginal cost equals price.

Of course, now that we have obtained the solution of the combined problem in terms of market prices p and $\mathbf{w}$ it would be interesting to know how the solution might be affected if those prices were to change.

2.3 The firm as a "black box"

We shall now see how we can put the firm's cost function to work: we use it to characterise the equilibrium of the firm in a simple way, and to analyse how the profit-maximising firm will react to changes in its market environment. We can imagine the firm to be like an electronic black box that accepts incoming signals from the market in the form of prices and, as the result of some predetermined inner workings, processes them and emits other signals in the form of quantities

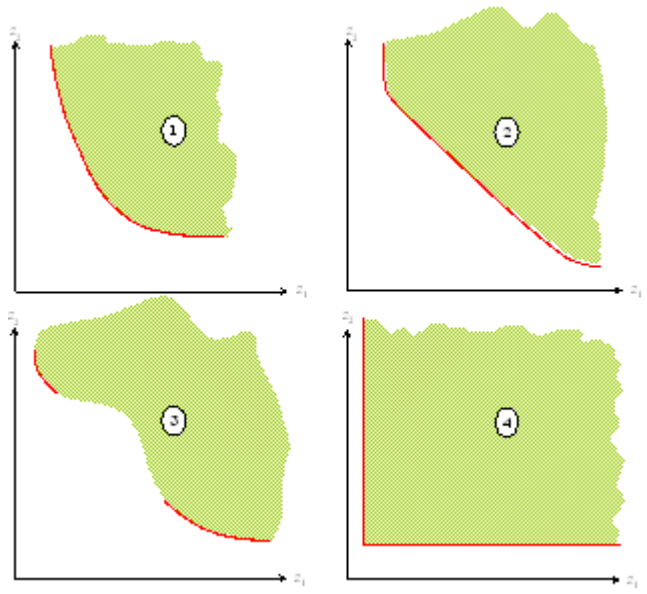

Figure 2.13: Convexity and input demands

of input demands and output supply. Our task is to characterise the inner workings of the black box.

To do this we shall make use of the method of *comparative statics*, which basically means that we see how the solution to the optimisation problem would change if some of the market data were to be altered a little. This can do more than just provide a simple mechanical response; it can reveal information about the structure of the solution as well. We shall then extend our analysis of the elementary model of the firm to cover two important developments: its reaction to "short-run" constraints, and the possibility of acting as a price-maker rather than as a price taker.

As a simple example of the basic comparative statics method let us go back to a point that we made earlier, that the nature of the solution to the firm's cost-minimisation problem would depend on the shape of the input-requirement set $Z(q)$. To examine the implications of alternative possible shapes for $Z(q)$ try the following four-part experiment:

- Take the case where $Z(q)$ is *strictly convex* – case 1 in Figure 2.13 – and use a straight-edge to represent the isocost line on the figure. Then, on a separate piece of paper, plot the cost-minimising value of z_1 against w_1/w_2, the slope of the isocost line; you should get a continuous, downward sloping curve. The shading of the boundary indicates the optimal **z**-values that you pick up as you do the experiment.

- If you conduct the same experiment for the case where $Z(q)$ is convex but

not strictly convex (case 2) you should find that you get a similar graph, but that there will be at least one point at which a single value of w_1/w_2 corresponds to an *interval* of values of z_1.

- Thirdly, try it for the case where $Z(q)$ is non-convex (case 3): you should find a point at which a single value of w_1/w_2 corresponds to exactly two values of z_1: between these two z_1-values there is a discontinuity in the relationship you are plotting.

- Fourthly, try it for the "kinked case": you will find that at a kink there is a range of w_1/w_2-values for which the optimal z_1 remains unchanged. However, although there is a unique input demand for a given w_1/w_2 value at the kinks, you will find a range of (w_1/w_2)-values which yield the same input demand.[19]

It is useful to compare Figures 2.1 and 2.13: note that not all the boundary points of $Z(q)$ (Figure 2.1) emerge as possible solution points in the cost-minimisation exercise (Figure 2.13) if $Z(q)$ is nonconvex. The experiment shows that the issue of convexity of the input-requirement set is central to the relationship between market prices and input demands. Also kinkedness of the boundary may destroy the uniqueness of the relationship between input demands and input prices. This will be put on a more formal basis in a moment; we will find that these insights apply in other aspects of economic optimisation.

2.3.1 Demand and supply functions of the firm

Let us follow up the point that emerged from the experiment, that for a suitably shaped $Z(q)$ – in other words a "well-behaved" production function ϕ – you would get a one-to-one relationship between the input price ratio and the demand for an input, but that for other production functions multiple solutions might emerge. This point – proved in Appendix C – is summarised more formally as:

Theorem 2.4 (Firm's demand and supply functions) *(a) If all input-requirement sets are strictly convex, conditional input demand functions are always well defined and continuous for all positive input prices. (b) If the production function is strictly concave, the supply function and input demand functions are always well defined and continuous for all positive input prices.*

The conditions required for the second half of this result are rather demanding. To see why this is so let us recall that the "conventional" supply relationship that we sketched in Figure 2.12 does not actually satisfy the requirements of part (b). If the average cost curve is U-shaped then the firm's supply of output is in fact multi-valued at one point: this is point $\underline{q}$, where p equals minimum

[19] Draw a case where $Z(q)$ is strictly convex *and* for which the boundary has multiple kinks. Draw the relationship between input price and conditional input demand and check that input demands are always uniquely defined.

average cost (given $p = \underline{p}$ the firm does not care whether it produces at $\underline{q}$ or produces nothing at all because it makes zero profits either way). This means that, strictly speaking, we have a supply *correspondence* rather than a supply function (see page 487 in Appendix A for this important technical distinction). The firm's supply curve is discontinuous at $\underline{p}$: there is a jump from 0 to $\underline{q}$ as the market price increases from a level just below $\underline{p}$ to just above $\underline{p}$. The reason for this is simple: the left-hand branch of the U-shape (to the left of $\underline{q}$) is a region where there is increasing returns to scale: the production function is not concave in this region.

Having thought about this, let us promptly ignore it for the moment and introduce three key concepts that we shall use frequently from now on. The first two are:

Definition 2.6 *The* conditional demand functions *for inputs $i = 1, 2, ..., m$ is a set of real-valued functions H^i of input prices and an output level such that*

$$z_i^* = H^i(\mathbf{w}, q) \tag{2.29}$$

where $(z_1^*, z_2^*, ..., z_m^*)$ are the cost-minimising inputs for $\mathbf{w}$ and q.

Definition 2.7 *The* supply function *of the competitive firm is a real-valued function S of prices such that*

$$q^* = S(\mathbf{w}, p) \tag{2.30}$$

where q^* is the profit-maximising output for $\mathbf{w}$ and p.

Notice that H^i must be homogeneous of degree zero in input prices $\mathbf{w}$, and that S is homogeneous of degree zero in $(\mathbf{w}, p)$.[20] Next, stick together these two principal solution functions that we have introduced. This then gives us the third key concept:

Definition 2.8 *The* unconditional demand function *for input i is a real-valued function D^i of input prices and the output price such that:*

$$z_i^* = D^i(\mathbf{w}, p), \tag{2.31}$$

where

$$D^i(\mathbf{w}, p) := H^i(\mathbf{w}, S(\mathbf{w}, p)). \tag{2.32}$$

Equation (2.32) emphasises that conditional and unconditional demands are just two different ways of tying down the same basic concept: in the first case we write the solution to the input-optimisation problem as function of input prices and output; in the second we write it as a function of input prices and the output price. Both versions are useful, as we shall see.

[20] Use the properties of the cost function to explain why this is so.

2.3.2 Comparative statics: the general case

Working with the supply curve is a simple example of comparative statics: we can show how q^* responds to p given the assumption of profit maximisation.

Suppose that we are in the interesting part of the problem where the firm is producing a strictly positive output. Then the points on the supply curve must also satisfy the standard first-order condition "price = marginal cost". Substituting in for q^* from (2.30), we may thus write:

$$p = C_q(\mathbf{w}, S(\mathbf{w}, p)). \tag{2.33}$$

where we have again used the subscript notation to represent the partial derivative. Differentiate (2.33) with respect to p and rearrange it to get:[21]

$$S_p(\mathbf{w}, p) = \frac{1}{C_{qq}(\mathbf{w}, q^*)}. \tag{2.34}$$

The left-hand side of (2.34) is the slope of the supply curve. The right-hand side depends on the way marginal cost C_q increases with output q. Since we know from the second order conditions that C_{qq} must be positive at the optimum, we see immediately from this that the competitive firm must have a rising supply curve.

Now consider input demands using the same sort of approach. Suppose the market price of output rises: as we know, output goes up, but what happens to input usage? Will a shift in the demand for the product also increase demand for, say, labour? Let us use the fundamental relationship between the two ways of writing input demands given in equation (2.32). Differentiating (2.32) with respect to p we get

$$D^i_p(\mathbf{w}, p) = H^i_q(\mathbf{w}, q^*)S_p(\mathbf{w}, p). \tag{2.35}$$

So the answer to our question is not quite straightforward: a rise in p will increase the demand for labour if and only if the term H^i_q is positive: this term is an "output effect" describing what would happen to conditional input demand if the specified level of output q were to be increased; the conventional assumption is that it is positive, so that z^*_i would go up as output level is increased (a "normal input"); but there are odd cases (so-called inferior inputs) where this does not happen. We can get further insight on this if we use Shephard's Lemma which, using (2.21) and (2.29), we may write as:

$$C_i(\mathbf{w}, q) = H^i(\mathbf{w}, q). \tag{2.36}$$

Then we find that (2.35) can be rewritten[22]

$$D^i_p(\mathbf{w}, p) = \frac{\partial C_q(\mathbf{w}, q^*)}{\partial w_i} S_p(\mathbf{w}, p). \tag{2.37}$$

[21] Do the differentiation and show this. You may find a review of the "function of a function" rule helpful – see section A.4.3.

[22] Show this, using the basic theorem on the properties of the cost function and the fact that the second partial differentials of C commute.

So, if the cost structure is such that an increase in the wage rate would have raised marginal cost, then we may deduce that an increase in product price would increase the employment of labour.

Now, what would happen to the demand for input i if the market price of input j were to alter? If the cost of paper (w_j) goes up do you employ fewer secretaries (z_i^*)? To address this issue, differentiate equation (2.32) again, this time with respect to w_j:

$$D_j^i(\mathbf{w}, p) = H_j^i(\mathbf{w}, q^*) + H_q^i(\mathbf{w}, q^*)S_j(\mathbf{w}, p). \tag{2.38}$$

We can simplify the second term on the right-hand side of this expression using the same sort of tricks as we have employed for earlier comparative-statics exercises. Using Shephard's Lemma the term H_q^i can be put in terms of the second derivative of the cost function; and differentiating (2.33) with respect to w_j we can get an expression for the required derivative of the supply function.[23] Substituting into (2.38) we find:

$$D_j^i(\mathbf{w}, p) = H_j^i(\mathbf{w}, q^*) - \frac{C_{iq}(\mathbf{w}, q^*)C_{jq}(\mathbf{w}, q^*)}{C_{qq}(\mathbf{w}, q^*)} \tag{2.39}$$

This fundamental decomposition formula for the effect of a price change can be expressed as follows:

$$\boxed{\begin{array}{c}\text{total}\\ \\ \text{effect}\end{array}} = \boxed{\begin{array}{c}\text{substitution}\\ \\ \text{effect}\end{array}} + \boxed{\begin{array}{c}\text{output}\\ \\ \text{effect}\end{array}}$$

The first component, the substitution effect, is the response that a firm would make to the input-price change if it were constrained to meet a fixed output target. The second component, the output effect, gives the change in input demand that is induced by a change in optimal output. Two nice results follow from the decomposition formula (2.39).

First, consider the substitution term H_j^i. Because of (2.36) we can write this term as C_{ij}, the cross-partial derivative of the cost function; and because $C_{ij} = C_{ji}$ (if the function is well-behaved, the order of differentiation does not matter) we see immediately that $H_j^i = H_i^j$ wherever the derivatives are well-defined. In other words all the substitution terms must be symmetric.

Second, have a look at the output effect term in (2.39). Clearly this too is symmetric in i and j. So since both this and the substitution term are symmetric we must also have $D_j^i = D_i^j$ for the uncompensated demands too: the overall cross-price effects are symmetric. So a rise in the price of paper would have the same effect on the (ordinary) demand for secretarial hours as would a rise in the wages of secretaries on the demand for paper.

Now let us think about the important special case where goods i and j happen to be the same, in other words the demand-response of input i to its own

[23] Do all this and derive (2.39).

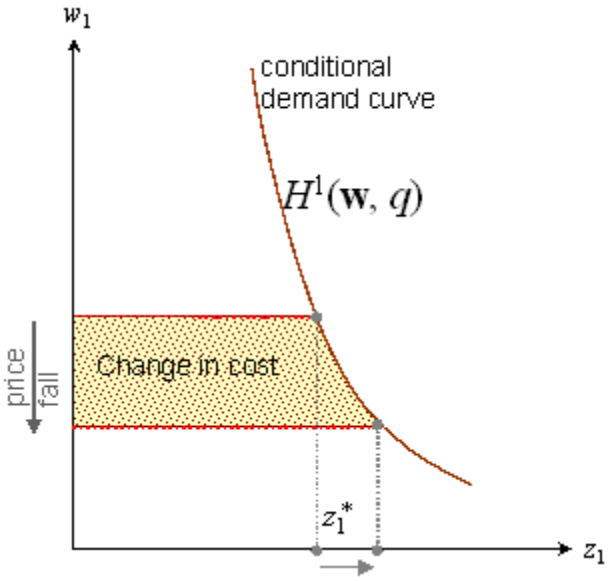

Figure 2.14: The substitution effect of a fall in price

price, w_i. Because C is concave in $\mathbf{w}$, we must have $C_{ii} \leq 0$ and hence $H_i^i \leq 0$.[24] In fact we can show that if ϕ were everywhere smooth and concave-contoured then, for all strictly positive input price vectors, we would have $H_i^i < 0$: the conditional demand for input i must be a decreasing function of its own price. Furthermore a quick check on the decomposition formula (2.39) reveals that in the own-price case we have:

$$D_i^i(\mathbf{w}, p) = H_i^i(\mathbf{w}, q^*) - \frac{C_{iq}(\mathbf{w}, q^*)^2}{C_{qq}(\mathbf{w}, q^*)} \tag{2.40}$$

We have just seen that the substitution effect in (2.40) is negative; so too, evidently, is the output effect (the squared term and C_{qq} are both positive); hence we have $D_i^i(\mathbf{w}, p) \leq 0$.[25]

We can pull all this together in the following statement:

Theorem 2.5 (Input prices and demands) *(a) The effect of an increase in the price of input j on the conditional demand for input i equals the effect of an increase in the price of input i on the conditional demand for input j; (b) the same result holds for the unconditional input demands; (c) the effect of an increase in the price of input i on the conditional demand for input i must*

[24] (For the mathematically inclined). Show this by using the result that a differentiable concave function must have a negative-semidefinite matrix of second partial derivatives – see page 507.

[25] Will the downward-sloping demand-curve also apply to consumer demand?

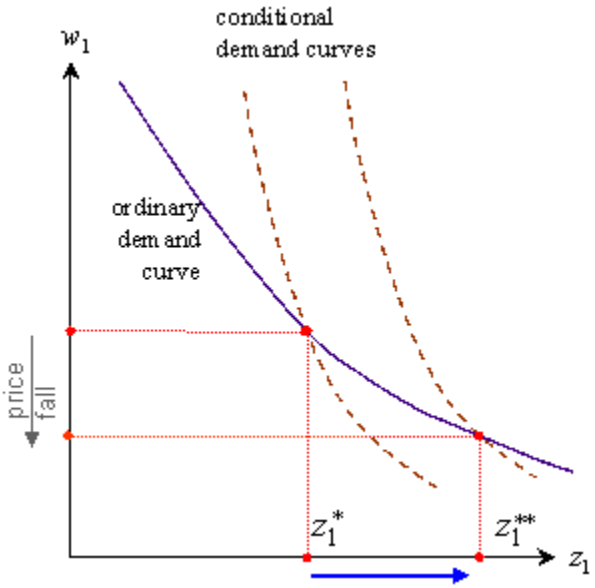

Figure 2.15: Input-price fall: total effect

be non-positive; (d) the effect of an increase in the price of input i on the unconditional demand for input i must be non-positive and greater in absolute size than the effect in (c).

We can use this information to sketch the shape of the demand curves for an input: – Figure 2.14 depicts the demand for input 1, conditional on a particular output level q. It must be downward sloping, because $H_1^1 < 0$ (Theorem 2.5). We also know that $H^1(\mathbf{w}, q)$ gives the marginal change in cost $C(\mathbf{w}, q)$ as w_1 changes (Shephard's Lemma): so the change in cost (for a fixed output q) resulting from a change in w_1 is given by the integral of H^1, which is depicted by the shaded area in Figure 2.14.

Let us consider the full effect of such a fall in w_1 such as that shown in Figure 2.14. It is obvious from Figure 2.14 that z_1 must increase, but that is purely a substitution effect. As we saw in equation (2.40) there is also an output effect; let us suppose that as w_1 falls the marginal cost curve in Figure 2.12 shifts downward so that output rises (the case of a normal input):[26] then the output effect is obviously positive, so that the total impact of the fall in input price is as shown in Figure 2.15.

Finally there is in this diagram a separate conditional demand curve for each level of output: that is why *two* conditional demand curves are drawn in – one for q^* (the original output level) and one for q^{**} (the output level after the price

[26] Notice that this reasoning implies that, for normal inputs, the ordinary demand curve is flatter than the conditional demand curve. Does the same apply to inferior inputs?

LONG RUN	
C	Cost function
D^i	Unconditional demand for input i
H^i	Conditional demand for input i
S	Supply function
SHORT RUN	
$\tilde{C}$	Cost function
$\tilde{H}^i$	Conditional demand for input i

Table 2.2: The Firm: Solution Functions

fall).

2.4 The short run

The short run is a notional period in which one or more inputs are assumed to be fixed. We introduce it to our model by taking input m to be fixed in the short run although, of course, it is variable in the long run.

- *Example 1: Capital Equipment.*[27] Take input m to be a mainframe computer. At some stage the firm has to decide how large a computer to install. The short-run curves are then derived on the assumption of a given size of computer, varying other inputs such as programmers' hours, secretarial hours, consumables..

- *Example 2: Employment protection.* Some types of workers may be able to negotiate long-term contracts with an employer. This section of staff in effect becomes a quasi-fixed factor.

To see the impact of this short-run fixity of an input, think of the behaviour of the profit-maximising firm as an mechanism, converting market data (prices) into supplies of output and demands for inputs. We have seen how this mechanism works in the comparative statics manipulations that we performed earlier on. Now suppose you tie down part of the system by imposing short-run constraints: what would we expect to happen? Presumably this will make the mechanism more sluggish – it will be less flexible in its response to changes in the market environment. This is in fact exactly what occurs.

To see this, let us introduce a proper definition of what we mean by the short run. Suppose that the conventional cost-minimisation problem has been solved for some specified output level $\overline{q}$ by setting input demands to $\overline{z}_1, \overline{z}_2, ..., \overline{z}_m$. By

[27] In what way might this be useful in representing firms' activities in a macroeconomic model?

definition we have:

$$\left.\begin{array}{c} \overline{z}_1 = H^1(\mathbf{w}, \overline{q}), \\ \overline{z}_2 = H^2(\mathbf{w}, \overline{q}), \\ \cdots \quad \cdots \\ \overline{z}_m = H^m(\mathbf{w}, \overline{q}). \end{array}\right\} \tag{2.41}$$

Now suppose that the specified output level is changed to some other value of q, but that the firm is constrained to keep its usage of the mth input fixed. Clearly it may want to alter its usage of the remaining m-1 variables; we will find the following concept useful:

Definition 2.9 *The firm's* short-run cost function *is a real-valued function* $\tilde{C}$ *of input prices, the output level, and an amount of input* m *such that:*

$$\tilde{C}(\mathbf{w}, q, \overline{z}_m) := \min_{\{z_i \geq 0, \phi(\mathbf{z}) \geq q, z_m = \overline{z}_m\}} \sum_{i=1}^{m} w_i z_i \tag{2.42}$$

The idea of these short run costs is that they are the best that you can do given that you are committed to an input level of $\overline{z}_m$ for the mth input.[28] Check this definition, term by term, against the definition of the firm's cost function in (2.19); in fact this function inherits – with very simple modifications – most of the conventional cost function's properties. In particular we have:

$$\tilde{C}_i(\mathbf{w}, q, \overline{z}_m) = \tilde{H}^i(\mathbf{w}, q, \overline{z}_m), \tag{2.43}$$

where $\tilde{H}^i(i = 1, ...m - 1)$ is the short-run demand for input i, conditional on output q, which emerges from the solution of the problem in (2.42).

By definition of the cost function, we must have

$$\tilde{C}(\mathbf{w}, q, \overline{z}_m) \geq C(\mathbf{w}, q). \tag{2.44}$$

Dividing both sides of (2.44) by q, we see immediately that long-run average cost must be less than or equal to short-run average cost. Of course, exactly at the point $q = \overline{q}$ it is true that:

$$\tilde{C}(\mathbf{w}, \overline{q}, \overline{z}_m) = C(\mathbf{w}, \overline{q}). \tag{2.45}$$

and therefore, at this point, $\partial \tilde{C}(\mathbf{w}, \overline{q}, \overline{z}_m)/\partial \overline{z}_m = 0$.

Let us look at the behaviour of long-run and short-run costs. What would have happened were we to have started from a different output level $\overline{q}$? Use (2.41) to write (2.45) as

$$\tilde{C}(\mathbf{w}, \overline{q}, H^m(\mathbf{w}, \overline{q})) = C(\mathbf{w}, \overline{q}) \tag{2.46}$$

[28] It is sometimes convenient to work with the concepts of short-run variable costs (the first $m - 1$ terms of the sum in the above definition) and of *fixed costs*, which are simply $w_m \overline{z}_m$. Show that the results which follow also work for short-run variable costs, rather than $\tilde{C}$, as defined.

and then differentiate this with respect to $\overline{q}$ so as to obtain, on simplification:[29]

$$\tilde{C}_q(\mathbf{w}, \overline{q}, \overline{z}_m) = C_q(\mathbf{w}, \overline{q}), \tag{2.47}$$

Thus, when output is at the level for the fixed input level $\overline{z}_m$ is optimal, long-run marginal costs (C_q) equal short-run marginal costs ($\tilde{C}_q$). Hence at $\overline{q}$ the slope of the long-run average cost curve must equal the slope of the short-run average cost curve. Using the same general method we can differentiate (2.45) with respect to w_i so as to obtain

$$\tilde{C}_i(\mathbf{w}, \overline{q}, \overline{z}_m) = C_i(\mathbf{w}, \overline{q}), \tag{2.48}$$

which implies

$$\tilde{H}^i(\mathbf{w}, \overline{q}, \overline{z}_m) = H^i(\mathbf{w}, \overline{q}). \tag{2.49}$$

So, in the neighbourhood of $\overline{q}$, short-run and long-run conditional input demands are identical.

Now let us look at the second-order conditions. Using the conditional input demand function for input m (see equation (2.41) above) differentiate (2.47) with respect to $\overline{q}$:

$$\tilde{C}_{qq}(\mathbf{w}, \overline{q}, \overline{z}_m) + \tilde{C}_{q\overline{z}_m}(\mathbf{w}, \overline{q}, \overline{z}_m) H_q^m(\mathbf{w}, \overline{q}) = C_{qq}(\mathbf{w}, \overline{q}), \tag{2.50}$$

Rearranging (2.50) we get:[30]

$$C_{qq}(\mathbf{w}, \overline{q}) = \tilde{C}_{qq}(\mathbf{w}, \overline{q}, \overline{z}) + \frac{H_q^m(\mathbf{w}, \overline{q})^2}{H_m^m(\mathbf{w}, \overline{q})} \tag{2.51}$$

But we know that the own-price substitution effect H_m^m must be non-positive (and if the production function is smooth it must be strictly negative). Hence for a locally smooth production function we find:

$$C_{qq}(\mathbf{w}, \overline{q}) < \tilde{C}_{qq}(\mathbf{w}, \overline{q}, \overline{z}) \tag{2.52}$$

In other words short-run marginal cost is steeper than long-run marginal cost.

In like manner by differentiating (2.49) with respect to $w_i (i = 1, 2, ..., m-1)$ we can derive[31]

$$H_i^i(\mathbf{w}, \overline{q}) > \widetilde{H}_i^i(\mathbf{w}, \overline{q}, \overline{z}), \tag{2.53}$$

so that short-run input demand is less elastic (to its own price) than long-run input demand.

We can summarise the above results thus:

[29] Explain why $\partial \tilde{C} / \partial \overline{z}_m = 0$ at $q = \overline{q}$, and prove (2.47).

[30] Show this. Hint: substitute the conditional demand function for $\overline{z}_m$ in (2.47) and differentiate (2.47) with respect to w_m, noting that $\partial \tilde{C}_q / \partial w_m = 0$ [Why?]; you then find an expression for $\tilde{C}_{q\overline{z}_m}$ to substitute in (2.50).

[31] Show this by following through the same steps as for short-run marginal costs, using Shephard's Lemma and the fact that the second derivatives of C commute.

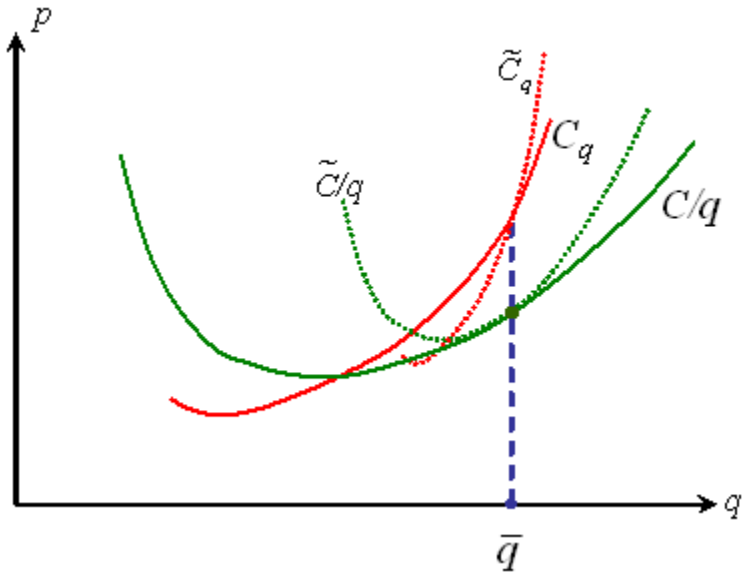

Figure 2.16: Marginal and average costs in the short and long run

Theorem 2.6 (Short-run demand and supply) *(a) Where output is at the optimal level for the fixed input, short-run and long-run total costs are equal. (b) At this output level, short- and long-run marginal costs are equal. (c) At this output level, short- and long-run input demands are equal. (d) The short-run marginal cost curve is at least as steep as the long-run marginal cost curve. (e) Long-run input demands are at least as elastic as short-run demands.*

Figure 2.16 illustrates these results in the case where long-run marginal costs are rising. Take the example where input m represents the computer the firm has just installed: technological change may have shifted the production function so that the firm now wishes it had a larger computer, but for now it is committed to the installation ($q = \overline{q}$). The broken cost curve represents the situation with the existing computer (allowing programmers' hours and materials to be varied in the short run);[32] the solid curve represents average costs given that computer installation can itself be taken as a variable input.

The results may be easily generalised. Instead of just one constraint, $z_m = \overline{z}_m$, let a further input be constrained, and then another and then another. Then we have the following for this sequential exercise:

$$\left.\frac{\partial z_i^*}{\partial w_i}\right|_{\text{no constraints}} \leq \left.\frac{\partial z_i^*}{\partial w_i}\right|_{\text{one constraint}} \leq \left.\frac{\partial z_i^*}{\partial w_i}\right|_{\text{two constraints}} \leq \ldots \qquad (2.54)$$

[32] Draw in on this diagram the short-run cost curves given that a computer system of ideal size had been installed.

a result which makes the "short run" as short as you like.

Example 2.1 *A classic study of US airlines (Eads et al. 1969) modelled long run costs as*

$$C(\mathbf{w}, q) = C_f + k' q^{\frac{1}{\gamma w_1}\frac{\alpha_1}{\gamma w_2}\frac{\alpha_2}{\gamma}} \tag{2.55}$$

where q is an index of airline output,C_f is the cost of fuel (separately estimated), w_1 is the price of labour other than pilots and copilots and w_2 is the price of pilots and copilots: the αs are parameters to be estimated econometrically, $\gamma = \alpha_1 + \alpha_2$, and k' is also a function of the αs. Differentiate (2.55) with respect to w_1 we get

$$z_i^* = H^1(\mathbf{w}, q) = \frac{\alpha_1 k}{\gamma} q^{\frac{1}{\gamma w_1}\frac{\alpha_{i-1}}{\gamma}\frac{\alpha_2}{\gamma}} \tag{2.56}$$

In other words the (long-run) conditional demand for labour of type 1 is given by the log-linear equation:

$$\log(z_i^*) = \beta_0 + \beta_1 \log(w_1) + \beta_2 \log(w_2) - \gamma \log(q) \tag{2.57}$$

Eads et al. (1969) assumed that in the short run pilots and copilots are a fixed factor (try sacking them!). The short-run cost function is then

$$\tilde{C}(\mathbf{w}, q, z_2) = C_f + k q^{\frac{1}{\alpha_1}} w_1 z_2^{-\alpha_2/\alpha_1} \tag{2.58}$$

Differentiating this with respect to w_1 we get the short-run demand for non-pilot labour which will also be log-linear. Try it.

2.5 The multiproduct firm

Clearly the assumption that the firm produces but a single output is rather limiting. To try to put this matter right we need another way of representing production possibilities. A method that is particularly convenient in the multiproduct case involves introducing one new concept – that of *net output*. Net outputs subsume both inputs and outputs using a natural sign convention under which outputs are measured in the positive direction ($q_i > 0$), and inputs negatively ($q_i < 0$).

Suppose there are n goods in the economy: the net output vector $\mathbf{q} := (q_1, ..., q_n)$ for the firm summarises all the firms' activities in the outside world. The firm's non-zero amounts of output or input for each good can be described according to the above sign convention; irrelevant goods, or pure intermediate goods can be ignored ($q_i = 0$). The production constraint[33] corresponding to (2.1) can be written

$$\Phi(\mathbf{q}) \leq 0 \tag{2.59}$$

where the function Φ is nondecreasing[34] in each of the q_i. A sectional snapshot of the multiproduct firm's production function is given in Figure 2.17: this

[33] Express the single-output production function (2.1) in this notation.

[34] Explain why it makes economic sense for Φ to be a non-decreasing function in each component, whether it be an input or an output.

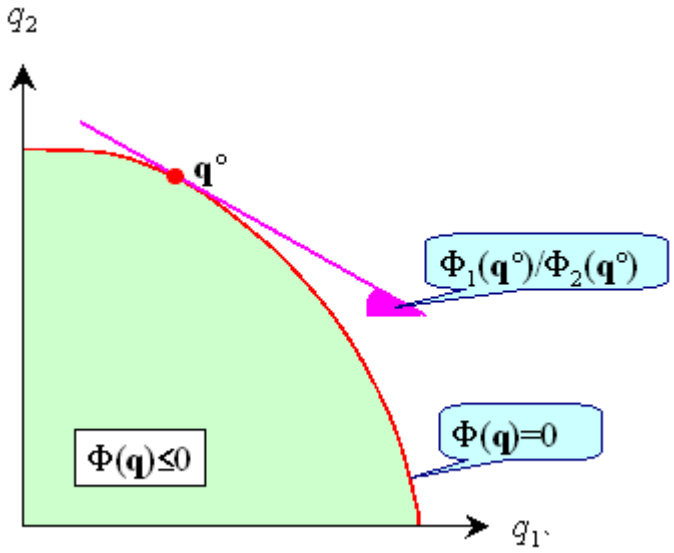

Figure 2.17: Firm's transformation curve

shows the production possibilities of two outputs that are potentially produced by the firm (of course the exact form of this snapshot depends on the values of the other components of the net-output vector – dimensions $3, 4, ..., n$). The shaded set depicts the net-output vectors that satisfy (2.59); the boundary of this set is known as the *transformation curve*.

There are obvious counterparts of assumptions about the single-output production function (see section 2.1.2 above) that can be easily established for Φ. Many of the standard concepts such as MRTS, marginal products and returns to scale[35] carry over straightforwardly to the multiproduct case: for the first two of these concepts the trick is usually to identify the appropriate contour of Φ. Obviously we have skated over these issues rather rapidly: we will have much more to say about them in chapter 6.

One important new concept can be defined wherever the production function is differentiable:

Definition 2.10 *The* marginal rate of transformation *of (net) output i into (net) output j is given by*

$$\text{MRT}_{ij} := \frac{\Phi_j(\mathbf{q})}{\Phi_i(\mathbf{q})}$$

[35] How would constant returns to scale be expressed in terms of the multi-output production function $\Phi(\cdot)$?

The MRT is the firm's trade-off or marginal valuation of a pair of goods – for example, the rate at which the firm would have to give up of one output in order to produce more of another. It has a central rôle to play in characterising market equilibrium (this is dealt with in chapters 6 and 7) and the efficiency of the allocation of goods and resources in an economy (chapter 9). Notice that the MRTS in definition 2.1 can be seen as a special case of definition 2.10 where goods i and j are both inputs.

One of the advantages of the net-output approach is that one has a particularly convenient expression for profits. To see this, imagine that for a particular firm the goods are labelled so that $1, ..., m$ are unambiguously inputs, goods $m+1, ..., r$ are are either intermediate goods or irrelevant, and goods $r+1, ..., n$ are unambiguously outputs (the labelling of goods is arbitrary, so we can always do this). The total value of inputs is given by:

$$\text{COST} = \sum_{i=1}^{m} p_i \left[-q_i\right] \tag{2.60}$$

where the term $-q_i$ is a positive number (because q_i is negative for inputs, under the convention); this is the absolute amount used of input i. The value of the outputs from the firm is obviously

$$\text{REVENUE} = \sum_{i=r+1}^{n} p_i q_i. \tag{2.61}$$

So, subtracting (2.60) from (2.61) and noting that the valuation of goods $m + 1, ..., r$ is zero (because here all the q_i values are zero) we find that

$$\text{PROFITS} = \sum_{i=1}^{n} p_i q_i. \tag{2.62}$$

The diagrammatic representation of profits works in just the same way as the diagrammatic representation of costs in Figure 2.10, but in the opposite direction – see the set of parallel *isoprofit* lines with slope $-p_1/p_2$ in Figure 2.18 that are the counterparts to the isocost lines in Figure 2.10. The firm's optimisation problem[36] then requires a solution to the constrained-maximum problem "maximise (2.62) subject to the feasibility condition (2.59)." Intuitively this involves reaching the highest isoprofit line in Figure 2.18 subject to remaining in the technologically feasible set (shaded in the figure). The method for solving this is in effect a modification of the cost-minimisation problem that we carried out for a fixed single output and a vector of m variable inputs in section 2.2.1. Formally we can represent this problem as that of maximising the Lagrangean:

$$\mathcal{L}(\mathbf{q}, \lambda; \mathbf{p}) := \sum_{i=1}^{n} p_i q_i - \lambda \Phi(\mathbf{q}) \tag{2.63}$$

[36] Re-express condition (2.27) for the multiproduct case.

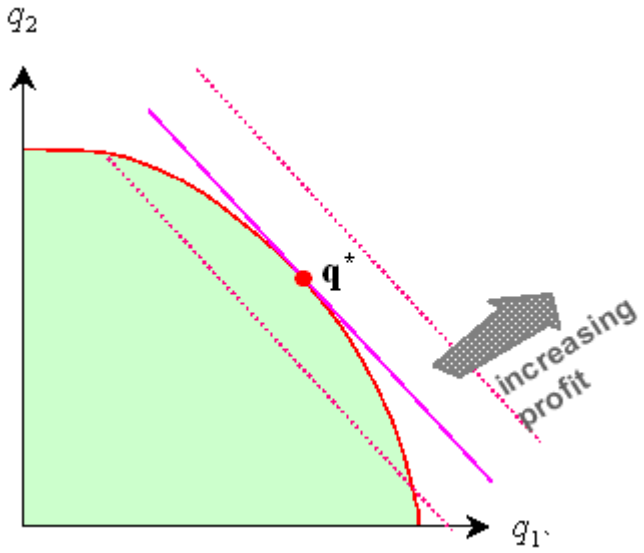

Figure 2.18: Profit maximisation: multiproduct firm

for given prices $\mathbf{p}$, where λ is the Lagrange multiplier associated with the constraint (2.59). Differentiating (2.63) with respect to q_i we can derive a set of first-order conditions that are the counterparts of the FOC in section 2.2.1. The result is a set of n profit-maximising net outputs $(q_1^*, ..., q_n^*)$ that satisfy the set of FOCs. In a manner similar to section 2.2.1 we then find:

- If net output i is produced in non-zero amounts at the optimum then

$$\lambda^* \Phi_i(\mathbf{q}^*) = p_i. \tag{2.64}$$

- For any pair of outputs i and j where output i is produced in positive amounts at the optimum the FOCs imply:

$$\frac{\Phi_j(\mathbf{q}^*)}{\Phi_i(\mathbf{q}^*)} \leq \frac{p_j}{p_i} \tag{2.65}$$

  with equality in (2.65) if input i is also used in positive amounts.[37]

- At the vector of optimal net outputs:

$$\Phi(\mathbf{q}^*) = 0. \tag{2.66}$$

[37] Draw a diagram to illustrate the case where "<" holds in (2.65). Give a brief verbal interpretation of the optimum.

So, once again we find that production is technically efficient and, in the case where profit-maximising amounts of both outputs are positive, we have the rule of thumb:

$$\boxed{\text{MRT}} = \boxed{\begin{array}{c}\text{output}\\ \text{price}\\ \text{ratio}\end{array}}.$$

Furthermore the result of this optimisation process is another solution function as follows:

Definition 2.11 *The firm's* profit function *is a real-valued function* Π *of net output prices such that:*

$$\Pi(\mathbf{p}) := \max_{\{\Phi(\mathbf{z})\leq 0\}} \sum_{i=1}^{n} p_i q_i \tag{2.67}$$

Clearly the profit function Π is the "twin" of the cost function C for the cost-minimisation problem in sections 2.2.1 and 2.2.2. So it is not surprising to find that there is a theorem characterising the properties of the profit function that is very similar to Theorem 2.2 for the cost function:

Theorem 2.7 (Properties of profit function) *The competitive firm's profit function* Π *is nondecreasing, continuous, homogeneous of degree one and concave in* $\mathbf{p}$*. At every point where the differential is defined*

$$\frac{\partial \Pi(\mathbf{p})}{\partial p_i} = q_i^* \tag{2.68}$$

the optimal value of net output i*.*

For proof see Appendix C. Equation (2.68) is usually known as Hotelling's Lemma and is established in the same way as Shephard's Lemma. In particular we can see that the part of the theorem about the slope of the profit function in equation (2.68) is obviously just Shephard's Lemma "turned around" in the case where i is an input. Other parts of the Theorem are proved in the same way as for Theorem 2.2.

We can push the analogy between the analysis of the multiproduct firm and the single product firm in sections 2.2 and 2.3 one stage further. Clearly the optimal net output value in (2.68) can be expressed as a function (or as a correspondence) of the price vector:

$$q_i^* = q_i(\mathbf{p}). \tag{2.69}$$

The properties of the net-output function $q_i(\cdot)$ in (2.69) follow from those of the single-output firm's demand and supply functions (see for example Theorems 2.4 and 2.5 and the associated discussion). So we find that $q_i(\cdot)$ is homogeneous of degree zero, is nondecreasing in its own price p_i and that, for any i and j:

$$\frac{\partial q_i(\mathbf{p})}{\partial p_j} = \frac{\partial q_j(\mathbf{p})}{\partial p_i}. \tag{2.70}$$

Clearly the analysis in terms of the profit function and net outputs has an attractive elegance. However it is not for the sake of elegance that we have introduced it on top of the more pedestrian output-as-a-function-of-input approach. We will find that this approach has special advantages when we come to model the economic system as a whole.

2.6 Summary

The elementary microeconomic model of the firm can be constructed rigorously and informatively with rather few ingredients. Perhaps the hardest part is to decide what the appropriate assumptions are that should be imposed on the production function that determines the firm's technological constraints.

The fundamental economic problem of the competitive firm can be usefully broken down into two subproblems: that of minimising the cost of inputs for a given output and that of finding the profit-maximising output, given that input combinations have already been optimally selected for each output level. Each of these subproblems gives rise to some intuitively appealing rules of thumb such as "MRTS = input price ratio" for the first subproblem and "price = marginal cost" for the second subproblem.

Changing the model by introducing side constraints enables us to derive a modified solution function (the short-run cost function) and a collection of modified response functions. We get the common-sense result that the more of these side constraints there are, the less flexible is the firm's response to changes in signals from the market.

The elementary model of the firm can usefully be generalised by what amounts to little more than a relabelling trick. Outputs and inputs are replaced by the concept of *net output*. This trick is an important step for the future development of the production model in chapters 6 and onwards.

2.7 Reading notes

On the mathematical modelling of production see Fuss and McFadden (1980). The classic references that introduced the cost function and the profit function are Hotelling (1932) and Shephard (1953). See also Samuelson (1983) chapters III and IV.

2.8 Exercises

2.1 *Suppose that a unit of output q can be produced by any of the following combinations of inputs*

$$\mathbf{z}^1 = \begin{bmatrix} 0.2 \\ 0.5 \end{bmatrix}, \mathbf{z}^2 = \begin{bmatrix} 0.3 \\ 0.2 \end{bmatrix}, \mathbf{z}^3 = \begin{bmatrix} 0.5 \\ 0.1 \end{bmatrix}$$

1. *Construct the isoquant for $q = 1$.*

2. *Assuming constant returns to scale, construct the isoquant for $q = 2$.*

3. *If the technique $\mathbf{z}^4 = [0.25, 0.5]$ were also available would it be included in the isoquant for $q = 1$?*

2.2 *A firm uses two inputs in the production of a single good. The input requirements per unit of output for a number of alternative techniques are given by the following table:*

Process	*1*	*2*	*3*	*4*	*5*	*6*
Input 1	*9*	*15*	*7*	*1*	*3*	*4*
Input 2	*4*	*2*	*6*	*10*	*9*	*7*

The firm has exactly 140 units of input 1 and 410 units of input 2 at its disposal.

1. *Discuss the concepts of technological and economic efficiency with reference to this example.*

2. *Describe the optimal production strategy for the firm.*

3. *Would the firm prefer 10 extra units of input 1 or 20 extra units of input 2?*

2.3 *Consider the following structure of the cost function:* $C(\mathbf{w}, 0) = 0, C_q(\mathbf{w}, q) = \text{int}(q)$ *where* $\text{int}(x)$ *is the smallest integer greater than or equal to* x*. Sketch total, average and marginal cost curves.*

2.4 *Draw the isoquants and find the cost function corresponding to each of the following production functions:*

$$\begin{aligned}
\text{Case A} &: q = z_1^{\alpha_1} z_2^{\alpha_2} \\
\text{Case B} &: q = \alpha_1 z_1 + \alpha_2 z_2 \\
\text{Case C} &: q = \alpha_1 z_1^2 + \alpha_2 z_2^2 \\
\text{Case D} &: q = \min\left\{\frac{z_1}{\alpha_1}, \frac{z_2}{\alpha_2}\right\}.
\end{aligned}$$

where q is output, z_1 and z_2 are inputs, α_1 and α_2 are positive constants. [Hint: think about cases D and B first; make good use of the diagrams to help you find minimum cost.]

1. *Explain what the returns to scale are in each of the above cases using the production function and then the cost function. [Hint: check the result on page 25 to verify your answers]*

2. *Discuss the elasticity of substitution and the conditional demand for inputs in each of the above cases.*

2.5 *Assume the production function*

$$\phi(\mathbf{z}) = \left[\alpha_1 z_1^{\beta} + \alpha_2 z_2^{\beta}\right]^{\frac{1}{\beta}}$$

where z_i is the quantity of input i and $\alpha_i \geq 0$, $-\infty < \beta \leq 1$ are parameters. This is an example of the CES (Constant Elasticity of Substitution) production function.

1. *Show that the elasticity of substitution is $\frac{1}{1-\beta}$.*
2. *Explain what happens to the form of the production function and the elasticity of substitution in each of the following three cases: $\beta \to -\infty$, $\beta \to 0$, $\beta \to 1$.*

2.6 *For a homothetic production function show that the cost function must be expressible in the form*

$$C(\mathbf{w}, q) = a(\mathbf{w})\, b(q).$$

2.7 *For the CES function in Exercise 2.5 find $H^1(\mathbf{w}, q)$, the conditional demand for good 1, for the case where $\beta \neq 0, 1$. Verify that it is decreasing in w_1 and homogeneous of degree 0 in (w_1,w_2).*

2.8 *Consider the production function*

$$q = \left[\alpha_1 z_1^{-1} + \alpha_2 z_2^{-1} + \alpha_3 z_3^{-1}\right]^{-1}$$

1. *Find the long-run cost function and sketch the long-run and short-run marginal and average cost curves and comment on their form.*
2. *Suppose input 3 is fixed in the short run. Repeat the analysis for the short-run case.*
3. *What is the elasticity of supply in the short and the long run?*

2.9 *A competitive firm's output q is determined by*

$$q = z_1^{\alpha_1} z_2^{\alpha_2} ... z_m^{\alpha_m}$$

where z_i is its usage of input i and $\alpha_i > 0$ is a parameter $i = 1, 2, ..., m$. Assume that in the short run only k of the m inputs are variable.

1. *Find the long-run average and marginal cost functions for this firm. Under what conditions will marginal cost rise with output?*
2. *Find the short-run marginal cost function.*
3. *Find the firm's short-run elasticity of supply. What would happen to this elasticity if k were reduced?*

2.10 *A firm produces goods 1 and 2 using goods 3,...,5 as inputs. The production of one unit of good i ($i = 1, 2$) requires at least a_{ij} units of good j, ($j = 3, 4, 5$).*

1. *Assuming constant returns to scale, how much of resource j will be needed to produce q_i units of commodity 1?*
2. *For given values of q_3, q_4, q_5 sketch the set of technologically feasible outputs of goods 1 and 2.*

2.11 *A firm produces goods 1 and 2 uses labour (good 3) as input subject to the production constraint*

$$[q_1]^2 + [q_2]^2 + Aq_3 \leq 0$$

where q_i is net output of good i and A is a positive constant. Draw the transformation curve for goods 1 and 2. What would happen to this transformation curve if the constant A had a larger value?