Chapter 5

Firm behaviour

5.1. Introduction

In this chapter we consider the second basic element of our theory: the firm. We assume that firms maximize profits, subject to production constraints. We start with the general case of a profit-maximizing firm with several inputs and outputs, and formulate some restrictions on the price elasticities following from the assumption of profit maximization. In section 3 we examine the case of perfect competition more closely, i.e. where excess profits are zero. It turns out that at the optimal point of production the production function exhibits constant returns to scale. The elasticities of substitution are introduced in section 4, where the consequences of profit maximization for these parameters are considered. In section 5 we attempt to reduce the number of unknown elasticities of substitution by specifying the production function as a CES production function. Under the assumption of cost shares and net prices being identical for all individual firms, we arrive in section 6 at supply relations for an aggregate of firms, which are fully compatible with the individual supply equations. In section 7 we present an alternative representation of aggregate firm behaviour, called the inverse supply function. As usual, we end with a summary of this chapter. As the reader will notice, parts of this chapter are almost identical to parts of chapter 4 concerning household behaviour. This is a direct consequence of the similarities in the objective functions of firms and households.
5.2. Micro-economic theory

The firm sector consists of a number of individual firms and it is to these individual firms that we shall turn now. As we did for the individual households, we examine firm behaviour in order to go behind the aggregate supply relation and see if it has to obey certain rules in order to be consistent with some kind of rational firm behaviour. Each firm can be seen as a black box where inputs come in one side and outputs come out the other side. The mechanism inside the black box is governed by the rule of profit maximization, given the technological possibilities as expressed by a production function. As mentioned in Chapter 2, firms are assumed to deal exclusively with private goods; public goods are not taken into account.

Concerning the time dimension, we notice two things. First, all inputs and outputs are measured as units per time interval or, in other words, as flows. This includes, of course, the amounts of services demanded; we speak, for example, of units of labour services demanded per year.

Secondly, concerning the time dimension of the production function, we make the assumption that only momentary flows enter it as arguments. Although this may seem to be a restrictive assumption, we do not think it is: we refer to Chapter 11 for a more extensive discussion of intertemporal problems. As a direct consequence of this assumption, there is no need to consider an intertemporal profit function, thus profit is also restricted to a function of momentary input and output flows. Firms are assumed to operate under certainty.

Now consider an individual firm, facing net prices $p_{Fn}$, where $n = 0, 1, \ldots, N$. The production function $f^*(\cdot)$ reflects the technological possibilities of transformation. It is nothing more than a function relating all inputs and outputs to each other. Net quantities supplied are denoted by $q_{Fn}, n = 0, 1, \ldots, N$. We assume there is at least one input and one output. Since this chapter deals exclusively with firms, the subscript F is dropped throughout (this holds also for variables not yet introduced, e.g., $c_n$ in (5.23) below stands for $c_{Fn}$, etc.). Thus, writing $q_n$ for $q_{Fn}$, we have

$$f^* (q_0, q_1, \ldots, q_N) = 0. \tag{5.1}$$

Notice that some quantities $q_n$ might be negative, representing negative net supply or, in other words, demand. Hence, positive quantities refer to outputs, while negative quantities refer to inputs. Using this convention, profit can be expressed by (with $p^*$ standing for $p_{F*}$)

$$\sum_{n=0}^{N} p_n q_n. \tag{5.2}$$

We will assume that all scarce factors used in production are explicitly considered as arguments of the production function, therefore appearing as cost factors into the profit function (5.2). Consequently, profits as defined in (5.2) correspond to excess profits, i.e., the profits above the total remuneration of scarce resources, the latter including the (sometimes implicit) payment for managerial and capital services. Notice that we assumed that firms are not subject to lump-sum taxes (or transfers): only transaction taxes, being related to demand and supply of goods, are relevant for firms.

The behaviour of the firm follows now from the firm's objective

$$\max \sum_{n=0}^{N} p_n q_n,$$

subject to

$$f^* (q_0, q_1, \ldots, q_N) = 0. \tag{5.3}$$

This problem has been examined extensively in the literature, just as was the analogous household problem [see eq. (4.5)].\(^1\) It is in fact closely related to the household problem, especially when there is only one output and several inputs. Here, we are concerned with the more general case of joint production as formulated above. We confine our discussion to the underlying assumptions and some straightforward derivations and results.

The main assumptions are: net prices are positive and given, exogenous variables, i.e., independent of the quantities supplied (this assumption rules out monopoly and monopsony power, and non-proportional taxes and discounts, see Chapter 11), while for a relevant region the production function $f^* (\cdot)$ is assumed to be continuously at least twice differentiable with positive first-order partial derivatives and negative semi-definite matrix of second-order partial derivatives. This last assumption concerning $f^* (\cdot)$ is sufficient (although not necessary) to guarantee a unique solution to the problem defined in (5.3).

\(^1\) See, for example, Intriligator (1971, ch. 8), and Malinvaud (1972, ch. 3).

\(^2\) A necessary and sufficient condition is that of quasi-concavity of the production function; in this case, the production set is convex. See Malinvaud (1972, p. 53).
\[ p_n = \psi \frac{\partial f^*}{\partial q_n}, \quad n = 0, 1, \ldots, N, \quad (5.4) \]

\[ f^*(q_0, q_1, \ldots, q_N) = 0, \quad (5.5) \]

assuming that all optimal quantities are feasible. \( \psi \) denotes the so-called Lagrange multiplier.

Given the above conditions, we obtain the solutions to the problem (5.3) as functions of the exogenous variables:

\[ q_n = q_n(p_0, p_1, \ldots, p_N), \quad n = 0, 1, \ldots, N. \quad (5.6) \]

If taxes or gross prices change, net prices \( p_n \) will in general also change. Since we assume tax changes to be small, we confine ourselves to the examination of marginal changes in prices. Subsequently, we will refer to the situation before the tax change as the initial situation. Being interested in comparative-static effects, we differentiate the net supply equations and reformulate the results in terms of elasticities. This yields

\[ \dot{q}_n = \sum_{n=0}^{N} \eta^*_n \dot{p}_m, \quad (5.7) \]

where

\[ \eta^*_{nm} = \frac{\partial q_n}{\partial p_m} q_n, \quad (5.8) \]

for \( n = 0, 1, \ldots, N \). The symbol \( \dot{\cdot} \) denotes relative changes, so \( \dot{x} \) equals \( dx/x \). The price elasticities \( \eta^*_{nm} \) are local parameters. They refer to the elasticities of the supply function (5.6) at the initial situation.

Now let us examine the implications of the theory for the price elasticities \( \eta^*_{nm} \). First, since profit is only scaled up when all net prices are changed proportionally, the solution (5.6) will not change either, given a production function independent of net prices. In other words, the supply functions are homogeneous of degree zero in net prices. So if \( \dot{p}_m = \alpha \), for all \( m \), with \( \alpha \) a small, non-zero number, we have

\[ 0 = \sum_m \eta^*_{nm} \alpha, \]

or

\[ \sum_{m=0}^{N} \eta^*_{nm} = 0, \quad n = 0, 1, \ldots, N. \quad (5.9) \]

Furthermore, total differentiation of the production function \( f^*(\cdot) \) yields

\[ \sum_{n=0}^{N} \frac{\partial f^*}{\partial q_n} \dot{q}_n = 0, \]

which becomes, using the first-order condition (5.4), in terms of relative changes

\[ \sum_{n=0}^{N} \epsilon_n \dot{q}_n = 0, \quad (5.10) \]

where the expenditures in the initial situation \( \epsilon_n \) are defined by

\[ \epsilon_n = -\frac{\partial u}{\partial q_n}. \quad (5.11) \]

These are positive for goods demanded and negative for goods supplied. Substituting (5.7) into (5.10) yields

\[ \sum_{n=0}^{N} \sum_{m=0}^{N} \epsilon_n \eta^*_{nm} \dot{p}_m = 0, \]

which should hold for all (small) values of \( \dot{p}_m \). Putting all \( \dot{p}_m \) except one equal to zero, we find

\[ \sum_{n=0}^{N} \epsilon_n \eta^*_{nm} = 0, \quad m = 0, 1, \ldots, N. \quad (5.12) \]

The restrictions (5.9) and (5.12) enable us to calculate the \( 2N + 1 \) numeraire elasticities \( \eta^*_{nm} \) and \( \eta^*_{mn} \) if we know all other price elasticities. Since there are \( 2N + 2 \) equations, it seems there is still one constraint left. This is not so. One of the equations of the system (5.9) and (5.12) can be derived from the others. For instance, the weighted sum (with weights \( \epsilon_n \)) over \( n \) of eq. (5.9) is identical to the sum over \( m \) of eq. (5.12). Thus, the above formulated results are just sufficient in number to express the numeraire elasticities as functions of the non-numeraire ones. Before considering the other constraints implied by the theory of profit maximization, we look somewhat more closely at the assumption of perfect competition.

5.3. Perfect competition

In the last section we assumed net prices to be given for the individual
firm. This assumption is one part of the assumption of perfect competition: the supply (and demand) of each individual firm constitutes such a small part of total supply (and demand) that the firm’s influence on prices is negligible. The other part of the assumption of perfect competition as adopted here, is called “free replication of production”. It has to do with the interaction between firms. Here, we leave the purely individual scene for a moment and focus upon a group of competing firms producing the same goods, called an industry.

The condition of “free replication of production” implies that production possibilities are open to everybody. Thus, new firms can enter the industry and produce in exactly the same way and at exactly the same costs as the firms already in the industry. Additionally, there are no barriers for firms to withdraw from the industry. In other words, for all competing firms in the industry (including potential ones), production possibilities (as reflected by the production function) and net prices are exactly the same. Notice that this excludes discriminatory taxation between firms within an industry. Under such conditions it is obvious that in the long run excess profits (defined as the profits above the total remuneration of scarce resources, the latter including the payment for managerial and capital services) must be zero. For if excess profits are positive, new firms would enter the industry, while in case of negative excess profits, i.e. losses, firms would leave the industry, thereby forcing market prices to their long-run equilibrium values where all competitive costs are just covered in equilibrium. Hence

$$\sum_{n=0}^{N} p_n q_n = 0. \quad (5.13)$$

This result holds, as mentioned above, only in the long run, provided that the assumption of free replication is valid. In the short run, firms may realize excess profits so long as competition from other firms does not appear. In case of imperfect competition, excess profits may exist even in the long run. Some consequences of imperfect competition will be discussed in Chapter 11. Here we focus upon the case of perfect competition, where (5.13) holds.

Since condition (5.13) holds in equilibrium before and after a change of taxes, we can differentiate it with respect to prices and quantities and express the result in terms of elasticities, i.e.

$$\sum_{s=0}^{N} c_s (\beta_s + q_s) = 0.$$  

Combining this result with eq. (5.10) we have

$$\sum_{n=0}^{N} c_n \beta_n = 0, \quad (5.14)$$

stating that the weighted average of relative net price changes, with weights equal to the corresponding initial expenditures, equals zero.

Furthermore, substituting the first-order condition (5.4) into the no-profit condition (5.13), we arrive at

$$\sum_{n=0}^{N} \frac{\partial f^*}{\partial q_n} q_n = 0. \quad (5.15)$$

This result can be converted (using Euler’s theorem$^3$) into the more lucid expression

$$\frac{\partial \ln f^* (\alpha q_0, \alpha q_1, \ldots, \alpha q_N)}{\partial \ln \alpha} = 0, \quad (5.16)$$

where the derivative is evaluated at $\alpha = 1$, i.e. at the point of the optimal quantities $q_0, q_1, \ldots, q_N$. This expression shows that under perfect competition, production is carried out at a point where locally constant returns to scale (CRS) holds: multiplying all quantities $q_i$ by a factor $\alpha$ (close to one) leaves the production $f^*$ unchanged, at least to a first-order approximation.$^4$ To illustrate this, consider the case $N = 1$ (two goods), where one good (labour services for example) is input and the other one is output. In fig. 5.1 we show the relation $f^* (q_0, q_1) = 0$ as well as the no-profit condition $p_0 q_0 = -p_1 q_1$. We chose output as numeraire, with index $n = 0$ (remember that input quantities are negative, so $q_1 < 0$). The no-profit condition implies that the optimal quantities $q_0$ and $-q_1$ should be situated on a ray through the origin with slope $p_1 / p_0$. That is,

$$q_0 = -\frac{p_1}{p_0} q_1.$$

Furthermore, the first-order conditions (5.4) imply

$^3$See, for example, Intriligator (1971, p. 467).
$^4$See also Kuenne (1963, p. 180).
mathematical complications, we will assume that the CRTS condition holds for a finite (although possibly limited) range: given an optimal set of quantities \( q_0, q_1, \ldots, q_N \), we have

\[
f^*(\alpha q_0, \alpha q_1, \ldots, \alpha q_N) = 0
\]

for all \( \alpha \) such that

\[
\alpha_0 < \alpha < \alpha_1, \quad \text{with } \alpha_3 < 1 < \alpha_1.
\]

Now, if \((q_0, q_1, \ldots, q_N)\) is (under perfect competition) a set of optimal quantities for given prices, \((\alpha q_0, \alpha q_1, \ldots, \alpha q_N)\) with \(\alpha_0 < \alpha < \alpha_1\) is also optimal since both the objective function (i.e. profit) and the constraint (i.e. \(f^* = 0\)) remain unchanged [see (5.13) and (5.17)]. So, for a given set of prices a whole range of optima results, i.e. there is no unique optimal solution. As a consequence, no supply function as formulated in (5.6) exists. Such a function is only defined if, for any set of prices, one and only one set of optimum quantities exists.

One way to avoid this dilemma is to select one good, with index \( s \), say, and to formulate the net supply relation conditionally upon its quantity \( q_s \). In this case, \( q_s \) fixes the scale of operation, while prices determine the quantity ratios, given the level of operation. The firm's objective function now becomes

\[
\max \sum_{n=0}^{N} p_n q_n,
\]

subject to

\[
f^*(q_0, \ldots, q_N) = 0,
\]

and

\[
q_s \text{ fixed.}
\]

This problem is completely analogous to the problem as specified in eq. (5.3). However, the dimension of the problem (i.e. the number of instruments) is now \( N \) instead of \( N + 1 \). Notice that, since \( q_s \) is fixed, \( p_s \) becomes irrelevant. Following the argument of section 2, we now express the supply relation, given \( q_s \), as
\[ a_n = a_n(p_0, \ldots, p_{n-1}, p_{n+1}, \ldots, p_N), \quad n \neq s, \quad (5.19) \]

or, in relative terms,
\[ \tilde{a}_n = \sum_{m=0}^{N} \eta_{nm} \tilde{p}_m, \quad n = 0, 1, \ldots, N, \quad (5.20) \]

where
\[ \eta_{nm} = \frac{\partial a_n}{\partial p_m} q_s, \quad n, m \neq s, \text{ keeping } q_s \text{ fixed}, \quad (5.21) \]
\[ \eta_{nm} = \eta_{mn} = 0, \quad n, m = 0, 1, \ldots, N. \quad (5.22) \]

The no-profit condition enables us to find the change in the price \( p_i \) of the scale variable. Defining the cost shares \( c_n \) with respect to the scale variable by
\[ c_n = -\frac{e_n}{e_s}, \quad (5.23) \]
where
\[ \sum_{s=0}^{N} c_s = 1, \quad (5.24) \]
we have, using (5.14),
\[ \sum_{s=0}^{N} c_s \tilde{p}_s = 0. \]

hence
\[ \tilde{p}_s = \sum_{n=0}^{N} c_n \tilde{p}_n. \quad (5.25) \]

The restrictions on the elasticities \( \eta_{nm} \) are also analogous to those on \( \eta^*_{nm} \). We have [see eqs. (5.9) and (5.12)]
\[ \sum_{m=0}^{N} \eta_{nm} = 0, \quad n = 0, 1, \ldots, N, \quad (5.26) \]

\[ \sum_{m=0}^{N} \eta_{nm} = 0, \quad m = 0, 1, \ldots, N. \quad (5.27) \]

As before, the two adding-up properties (5.26) and (5.27) are not mutually independent. One of the restrictions follows from the remaining ones. Eqs. (5.26) and (5.27) allow us to express the elasticities corresponding to the numeraire as functions of the non-numeraire ones.

### 5.4. Substitution and scale effects

If prices are held fixed, changing the scale quantity \( q_s \) induces proportional changes in all other quantities, in view of the CRTS property of the production function. Hence, we may extend eq. (5.20) to
\[ \tilde{q}_n = \sum_{m=0}^{N} \eta_{nm} \tilde{p}_m + \tilde{q}_s, \quad n = 0, 1, \ldots, N. \quad (5.28) \]

Inspection of these net supply equations in relative terms shows a similarity to the equations for the household case. Instead of the income effect, however, we notice a scale effect, where scale is related to \( q_s \). Analogously to the household case, we can also distinguish a substitution effect, holding \( q_s \) constant. Since there is no income effect as prices increase, the elasticities \( \eta_{nm} \) represent only the substitution effects. We define the elasticities of substitution \( \sigma_{nm} \) implicitly by
\[ \eta_{nm} = c_n \sigma_{nm}, \quad (5.29) \]
where
\[ \sigma_{ss} = \sigma_{sm} = 0, \quad n, m = 0, 1, \ldots, N. \quad (5.30) \]

As with the other parameters, these elasticities are local, i.e. describing the characteristics of the supply relation around the initial situation. They also depend upon the choice of \( s \).

From the theory of profit maximization, some restrictions follow for the elasticities of substitution \( \sigma_{nm} \) (and therefore for the price elasticities \( \eta_{nm} \)). We will state them here without proof; the interested reader is referred to the references.\(^5\) The restrictions are, for \( n,m = 0,1,\ldots,N \).

\(^5\)See footnote 1.
\[ \sigma_{nm} = \sigma_{mn}, \quad (5.31) \]

\[
\sum_{n=0}^{N} \sum_{m=0}^{N} \sigma_{n} \sigma_{nm} \alpha_{m} = 0, \quad \text{if } q_{m} > 0,
\]

\[
\geq 0, \quad \text{if } q_{m} < 0,
\]

(5.32)

for all real values of \( \sigma_{n} \). Without loss of generality, we assume that \( q_{n} \) is non-zero. The first condition, eq. (5.31), states that the substitution effect, measured in terms of the elasticity of substitution, is symmetric. This result has the practical advantage that it reduces the number of unknown elasticities by nearly one-half.

The second condition, eq. (5.32), implies that the matrix of elasticities of substitution be either negative semi-definite (if \( q_{n} \) is positive) or positive semi-definite (in the opposite case). Choosing the values of \( \alpha_{0}, \alpha_{1}, \ldots, \alpha_{N} \) such that they are all zero except for a single \( \alpha_{n} \), implies

\[
\sigma_{nn} \begin{cases} 
\leq 0, & \text{if } q_{n} > 0, \\
\geq 0, & \text{if } q_{n} < 0.
\end{cases}
\]

(5.33)

Finally, from (5.26) and (5.31) we find

\[
\sum_{n=0}^{N} \sigma_{n} \sigma_{nm} = \sum_{n=0}^{N} \sigma_{nm} \alpha_{m} = 0, \quad m = 0, 1, \ldots, N.
\]

(5.34)

Notice that, unlike the household case, the above restrictions enable us to concentrate on \( \frac{1}{2}N(N-1) \) unknown price elasticities: the adding-up conditions (5.26) and (5.27) eliminate \( 2N - 1 \) elasticities, while the symmetry condition (5.31) further reduces the remaining \( (N-1)(N-1) \) elasticities to \( \frac{1}{2}(N-1)N \). Given these restrictions on the elasticities, we describe the behaviour of the competitive firm by eq. (5.28), while competition restricts prices such that eq. (5.25) holds.

5.5. Reducing the number of unknowns

Assuming that all expenditures \( e_{n} \) (and therefore the cost shares \( c_{n} \)) are observable, we have \( \frac{1}{2}N(N-1) \) unknown elasticities of substitution. For a simple three-good model (output, labour, and capital, for example), \( N \) equals 2, so the total number of unknowns equals one. For \( N = 9 \) (a ten-good economy), the total number of unknowns climbs to 36. Al-

though these figures are less than the corresponding number of unknowns in the household case, it still remains a virtually impracticable job to estimate all these parameters when \( N \) is large. Therefore, we will devote this section to an attempt to reduce the number of unknown parameters. We will do so by assuming all cross-price elasticities of substitution to be equal. So we have, using eq. (5.34),

\[
\sigma_{nm} = \sigma, \quad n \neq m \quad \text{and} \quad n,m \neq s, \quad (5.35)
\]

\[
\sigma_{nn} = (1 - c_{s}^{-1}) \sigma, \quad n \neq s. \quad (5.36)
\]

In order to meet condition (5.33) we assume here either one output and several inputs or one input and several outputs. Given a particular choice of the scale variable (i.e. of the index \( s \)) we might reformulate the production function as

\[
q_{s} = f(q_{0}, q_{1}, \ldots, q_{s-1}, q_{s+1}, \ldots, q_{N}). \quad (5.37)
\]

If all cross elasticities of substitution are equal, the production function \( f(\cdot) \) could be approximated locally by a so-called constant elasticity of substitution (CES) production function. Analogously to the terminology for utility functions, we will sometimes speak of a one-level production function (multi-level production functions are introduced in the appendix to this chapter).

The elasticity \( \sigma \) has been subject to much econometric estimation. For the simple model of one output and two inputs, namely capital and labour, values around \( \sigma = 1 \) were often found if output was taken as the scale variable. In this particular case we have \( (n,m \neq s) \),

\[
\eta_{nn} = c_{n}, \quad n \neq m, \quad \eta_{nn} = c_{n}^{-1}
\]

and, using (5.21) and (5.25), we find for the relative change in cost shares \( c_{n} \), holding output constant,

"Recently, new "flexible" forms of production functions (able to provide a second-order approximation) have attracted some attention, especially since these forms seem to provide possibilities of testing special nesting (or separability) assumptions (see also the appendix to this chapter). The interested reader is referred to Berndt and Christensen (1973), Blackorby et al. (1977), Conrad and Jorgensen (1977), and Denny and Fuss (1977). "See, for some recent estimates, Berndt (1976)."
Thus, for the one-level production function, with one output (being the scale variable), a unitary cross-elasticity of substitution corresponds to some kind of benchmark: if prices change, quantities change in such a way as to keep cost shares constant. In this case the production function \( f(\cdot) \) could be locally approximated by a so-called Cobb-Douglas production function, which is a member of the CES class. If \( \sigma > 1 \), one can easily prove that cross-price effects on cost shares are positive (and own-price effects on cost shares negative). The opposite holds if \( \sigma < 1 \). Notice that conditions (5.33) place a restriction on \( \sigma \), namely (since \( q_s > 0 \))

\[
\sigma \geq 0. \tag{5.38}
\]

If \( \sigma = 0 \), the production function is of the Leontief type: holding output constant, prices do not influence input quantities and, as a result, quantity shares \( q_m/q_s \) remain constant.

Notice that the sign of the elasticity of substitution depends on the production structure. If there is only one output the elasticity of substitution \( \sigma \) must be non-negative, while if there is only one input, the elasticity of substitution must be non-positive, in view of condition (5.35) [hint: use eqs. (5.36) and (5.23)].

The assumption that all cross-elasticities of substitution are equal is of course a restrictive one (if \( N > 2 \), the more so according to whether the number of goods \( (N + 1) \) is larger. Then it becomes preferable to assume group-wise equality of cross-price elasticities of substitution, as we did for the household in the appendix to Chapter 4. This brings us to the so-called production tree or nested production function. Since this subject is somewhat outside the scope of the book, we refer to the appendix of this chapter for further details.

### 5.6. Aggregation over firms

Until now we have considered the behaviour of a single, individual firm. In this section we focus on aggregates of firms and see how our theories might be extended to cover such groups.

---

1. See also de Boer (1976).
2. If \( \sigma = 0 \), the isoquants are rectangular-shaped, giving rise to non-differentiable production functions. Optimal solutions to the profit-maximizing problem do exist, however, since in this case the production set is convex.
\[ \check{\rho}_n = \check{\rho}_n, \quad f = 1, \ldots, F. \] (5.45)

In other words, net prices are always equal for all firms. This assumption excludes differentiated taxes with respect to firms (think of subsidies). We might have relaxed this assumption in the sense that net prices for different firms are not identical but merely proportional to each other, but we prefer not to do so in order to keep things simple. The general case of differentiated firm taxes will be examined in Chapter 10.

Our second assumption is rather restrictive. We assume for the initial situation that, for \( n = 0, 1, \ldots, N \),

\[ c_n^f = c_n, \quad f = 1, \ldots, F, \] (5.46)

i.e. all initial cost shares are assumed to be identical for all firms. This assumption is much more restrictive than the corresponding assumption of equal marginal shares in the household case. Here, the structure of production in terms of expenditure ratios is fixed, leaving only scale differences and elasticities of substitution to discriminate between firms. Although we do not need the assumption of identical elasticities of substitution for our (local) results, the assumption of identical initial cost shares becomes rather unrealistic if substitution elasticities are different. In this case, the equality (5.46) represents a remarkable coincidence, since price changes will induce non-identical changes in cost shares for different firms. Consequently, the assumption of identical cost shares might be too restrictive for operational models of the aggregate firm sector. For this reason a disaggregated firm sector is examined in Chapter 10.

Substituting eq. (5.39) into eq. (5.41) and using eq. (5.44) yields

\[ q_n = \sum_j a_j^f \sum_m c_n^f \sigma_{nm}^f \check{\rho}_m + \sum_j a_j^f q_j, \]

which becomes, by means of eqs. (5.45) and (5.46),

\[ \check{q}_n = \sum_{m=0}^N n_m \sigma_{nm} \check{\rho}_m + \check{q}_n, \] (5.47)

where

\[ c_n = -\frac{\partial q_n}{\partial q_j} = c_n^f, \] (5.48)

\[ \sigma_{nm} = \sum_j a_j^f \sigma_{nm} = \sum_j \left( \frac{e_j}{e_j} \right) \sigma_{nm}, \] (5.49)

\[ e_n = -p_n q_n. \] (5.50)

Eq. (5.47) can be written as

\[ \check{q}_n = \sum_{m=0}^N n_m \check{\rho}_m + \check{q}_n, \] (5.51)

where

\[ n_m = c_n \sigma_{nm}, \quad n = 0, 1, \ldots, N. \] (5.52)

We see that, under the assumptions made, the structure of the aggregate supply relation (in relative terms) is in conformity with the individual relations. The aggregate elasticity of substitution \( \sigma_{nm} \) turns out to be a weighted average of the individual elasticities, with weights equal to the shares \( c_n \sigma_{nm} \). It is straightforward to prove that all usual restrictions on individual parameters also hold for the corresponding aggregate parameters. The same is true for the price restriction (5.40). On the aggregate level we have

\[ \sum_{n=0}^N c_n \check{\rho}_n = 0. \] (5.53)

Thus, the (marginal) behaviour of the aggregate firm sector, under the assumptions made, is not distinguishable from that of an individual firm. In other words, the behaviour of the aggregate firm might be represented by that of a particular firm, called the representative firm, with characteristics equal to aggregates of individual characteristics.

Since we are not interested in individual firms but in certain groups of firms (at least with regard to tax incidence), we will subsequently specify our firm sectors in terms of variables and parameters of the representative firm. In this case the specified (aggregate) elasticities of substitution \( \sigma_{nm} \) must be interpreted as weighted averages of the underlying individual elasticities.

Finally, some remarks are in order about aggregation and competition. The assumption of identical cost shares in the initial situation implies that each firm in the aggregate produces the same goods, using the same
factors. So, what we considered is, in fact, an industry. Consequently, technology and net prices are identical between firms, as a result of the assumption of “free replication” (see section 3 above). More precisely, if the individual production function exhibits CRTS globally, competing firms in an industry may differ only with respect to scale, while elasticities of substitution should be the same. Then, the above procedure amounts to the rather trivial exercise of aggregating firms which are identical, except possibly for scale of operation. Thus, while the assumptions made here seem much more restrictive than those made in the household case (see Chapter 4, Section 6), they are not necessarily less realistic, since perfect competition forces firms to use identical technologies, while no such implication follows for the household case. Of course, the validity of this reasoning is conditional upon the realism of the assumption of perfect competition.

5.7. The inverse supply function

As said above, we are not interested in individual firms, but (at least with regard to tax incidence) only in groups of firms, such as industries. In the last section we showed that the behaviour of the industry might be represented by eqs. (5.51) and (5.53). Instead of these two (sets of) equations, we often prefer to use the so-called inverse supply functions, which state the changes (in) prices as functions of (the changes in) quantities. To arrive at such relationships we proceed as follows. First, we delete the equation for the numeraire \( (n = 0) \) from the set of equations as given in (5.51). For the moment we assume that the numeraire is not the scale variable. Using the convention for the numeraire price,

\[
\bar{p}_0 = 0,
\]

we can also delete \( \bar{p}_0 \) from eqs. (5.51) and (5.53). The resulting set of equations reads,

\[
\bar{q}_n = \sum_{m=1}^{N} \eta_{nm} \bar{p}_m + \bar{q}_n, \quad n = 1, 2, \ldots, N,
\]

\[
\sum_{n=1}^{N} c_n \bar{q}_n = 0.
\]

In view of eq. (5.22), eq. (5.55) represents \( N - 1 \) independent equations in \( N - 1 \) quantity changes \( \bar{q}_n \) and \( N - 1 \) price changes \( \bar{p}_n, n = 1, 2, \ldots, s - 1, s + 1, \ldots, N \), while eq. (5.56) contains only the \( N \) price changes \( \bar{p}_n, n = 1, 2, \ldots, N \). Combining these equations, we have \( N \) equations in \( N \) price changes \( \bar{p}_n \), which under general conditions could be solved for \( \bar{p}_n \). The result is, in relative terms, the inverse supply function:

\[
\bar{p}_n = \sum_{m=1}^{N} \mu_{nm} \bar{q}_m, \quad n = 1, 2, \ldots, N.
\]

The inverse supply elasticities \( \mu_{nm} \) must obey some restrictions. First, under a proportional change in all quantities, prices remain unchanged, due to CRTS, so

\[
\sum_{n=1}^{N} \mu_{nm} = 0, \quad n = 1, 2, \ldots, N.
\]

Secondly, eq. (5.56) must hold for all (small) changes \( \bar{q}_m \), so

\[
\sum_{n=1}^{N} c_n \mu_{nm} = 0, \quad m = 1, 2, \ldots, N.
\]

These restrictions on \( \mu_{nm} \) show that the matrix \( M = [\mu_{nm}] \) is singular; under perfect competition we cannot solve for a supply function as given in eq. (5.7). Notice that, unlike the elasticities \( \eta_{nm} \), the elasticities \( \mu_{nm} \) are independent of the choice of the scale variable.

If the scale variable and the numeraire coincide, i.e. if \( s = 0 \), the scale quantity on the right-hand side of eq. (5.55) is eliminated by using eqs. (5.10) and (5.23). We have then

\[
\bar{q}_n = \bar{q}_0 = \sum_{m=1}^{N} c_n \bar{q}_m,
\]

so that eq. (5.55) becomes, for \( n = 1, \ldots, N, \)

\[
\bar{q}_n = \sum_{n=1}^{N} \eta_{nm} \bar{p}_m + \sum_{n=1}^{N} c_n \bar{q}_m.
\]

Using eq. (5.27) we can easily show that these equations are not
mutually independent: one of the equations might be deleted. Together with (5.56) we have again \(N\) equations in \(N\) price changes \(\bar{p}_n\), which can be solved to obtain the inverse supply function as given in eq. (5.57).

Finally, it might be instructive to see what the inverse supply function looks like for the one-level production function. For simplicity we assume the scale variable to be the numeraire. By means of (5.23), (5.35), (5.36), and (5.56), (5.60) now becomes

\[
\bar{q}_n = -\sigma \bar{p}_n + \sum_{m=1}^{N} c_m \bar{q}_m, \quad n = 1, 2, \ldots, N.
\]

Hence, the inverse supply function reads

\[
\bar{p}_n = -\sigma^{-1} \left( \bar{q}_n - \sum_{m=1}^{N} c_m \bar{q}_m \right), \quad n = 1, 2, \ldots, N.
\] (5.61)

If \(N = 2\), one might use \(c_1 + c_2 = 1\) and write these equations in matrix form as

\[
\begin{bmatrix}
\bar{p}_1 \\
\bar{p}_2 
\end{bmatrix} = \sigma^{-1} \begin{bmatrix}
-c_2 & c_2 \\
c_1 & -c_1
\end{bmatrix} \begin{bmatrix}
\bar{q}_1 \\
\bar{q}_2 
\end{bmatrix}.
\] (5.62)

Since there is only one relevant cross elasticity of substitution if \(N = 2\), eq. (5.62) provides the general representation in this case.

5.8. Summary

In this chapter the behaviour of the firm is considered. We assume that firms maximize profits, subject to production constraints. We start with the general case of a profit-maximizing firm with several inputs and outputs. Assuming net prices to be given for each firm, we arrive at net supply functions with these prices as arguments. Negative net supplies correspond to demands. As usual, we focus on the effects of small changes in exogenous variables. The price elasticities of net supplies are now of special interest. As for the household, we show that these elasticities should obey some adding-up restrictions, which enables us to determine the elasticities corresponding to the numeraire if all other elasticities are known.

The assumption of perfect competition as adopted here consists of two parts. First, (net) prices are exogenous to the individual firm. Secondly, there is free replication within an industry: everybody can produce in exactly the same way and at exactly the same costs. Under this assumption competition between firms forces excess profits to be zero in the long run. Excess profits are defined as the profits above the total remuneration of scarce resources, where the payments for managerial and capital services are assumed to be included in the latter. We demonstrate that, under the assumption of perfect competition, production is carried out at a point where (locally) constant returns to scale (CRTS) holds. In this case it can be shown that net supply functions in terms of prices are undefined, since the optimal level of production is not unique. This problem is circumvented by choosing one quantity as a scale index and formulating the net supply functions conditionally on this quantity.

So we distinguish two effects on net supplies: the scale effect and the effect of prices. The elasticities corresponding to the price effects are directly related to the elasticities of substitution. From the theory of profit maximization we know that the matrix of these elasticities is semi-definite and symmetric. The last restriction is of practical interest since it reduces the number of unknowns by nearly one-half. In section 5 and in the appendix to this chapter some further attempts are made to reduce this number by specifying the production function as a (nested) CES function.

Under the assumption of cost shares and net prices being identical for all individual firms, we arrive at net supply functions for an aggregate of firms which are fully compatible with the individual supply functions. Therefore, the (marginal) behaviour of an aggregate of firms might be represented by that of a representative firm with characteristics equal to aggregates of individual characteristics. To describe the behaviour of this firm, one might use the inverse supply function instead of the ordinary supply function, since the former is invariant with respect to the choice of the scale variable, in contrast to the latter.

Appendix: A nested-CES production function

In this appendix we consider alternative ways of reducing the number of unknown elasticities. We do so under the additional assumption of absence of joint production, i.e. existence of a single output only. Output is also chosen as the scale variable.
A nested production function is a special form of the production function, \( f(\cdot) \), where output is defined as a function of aggregates of inputs. These aggregates are in turn defined as functions of aggregates at a lower level and so on. Consider the example where output depends on value-added and materials. Value-added is in turn a function of capital and labour services, which are two inputs. It will be clear that this kind of “nesting” can be done for all arguments of the production function. In fig. 5.2 the decomposition of aggregates into lower level aggregates is illustrated. Let us now try to formalize the idea.10

The nested production function consists of \( L + 1 \) levels \( l = 0, 1, \ldots, L \). At each level we distinguish several production components. At the highest level (indicated by \( l = L \)) of the production tree there is only one component, which coincides with output. This component is a function of production components at the next-lower level, \( l = L - 1 \). These production components at level \( L - 1 \) are in turn each a function of separable groups of production components at level \( L - 2 \), and so on. Finally, the production components at level \( l = 1 \) are functions of the \( N \) production components at level \( l = 0 \), which we identify with the \( N \) quantities of inputs.

Each production component at level \( l = 1, \ldots, L \) represents the “production” of an aggregate of some production components at lower levels. The following definition will prove useful: we say that two production components at different levels are associated if the higher level component is a function of the component at the lower level. This means that two components are associated if one component is an aggregate that includes the other. Consequently, in the example given above, the components labour, value-added, and output are associated: labour is included in value-added and value-added is included in output.

Since “value-added” is a composite, a physical quantity measure for value-added does not exist (we cannot add quantities of different items together). Like the other aggregates, the component “value-added” is just an index representing the contribution to production.

We now introduce the following notation: we write \( q_{n,l} \) for the production component at level \( l \) associated with input \( q_n \). Thus, in our example, if \( q_n \) is defined as the quantity of labour services, \( q_{n,1} \) corresponds to value-added and \( q_{n,2} \) to output. Notice that for a particular production component (except those at level zero) the notation \( q_{n,j} \) is not necessarily unique. If \( q_n \) is the quantity of labour services and \( q_n \) the quantity of capital services, \( q_{n,1} \) equals \( q_{n,1} \); both refer to the component value-added.

By definition, \( q_{n,l} \) and \( q_{n,k} \) are associated if the higher level component, say \( q_{k,l} \) (for \( l \leq k \)) is an aggregate which includes the lower level component \( q_{n,k} \). In other words, \( q_{n,k} \) “belongs” to the aggregate \( q_{n,l} \) and therefore we introduce the notation

\[
q_{n,k} \in q_{n,l},
\]

if \( q_{n,k} \) and \( q_{n,l} \) are associated and \( k \leq l \). Now, the nested production function can be formulated as follows (remember that input \( q_n \) is measured negatively):

\[
q_{n,0} = - q_n, \quad \text{(5.63)}
\]

\[
q_{n,l} = q_n(q_{n,l-1}; m \in n), \quad l = 1, \ldots, L, \quad \text{(5.64)}
\]

\[
q_{n,L} = q_{n}, \quad \text{(5.65)}
\]

for \( n = 0, 1, \ldots, s - 1, s + 1, \ldots, N \), using the notation \( m \in n \) to indicate

---

9The idea of a nested production function is not new. A global two-level CES production function was introduced by Satz (1967). Nesting implies separability; see footnote 6. The subsequent analysis bears heavily on Keller (1976).
those \( m \) for which \( q_{m,l-1} \in q_{n,l} \). An example of the structure and notation of the nested production function is given in fig. 5.3.

Having defined the nested production function, we are now able to use the concept in order to reduce the number of unknown elasticities of substitution. Instead of assuming all cross-price elasticities of substitution to be equal, we assume that they are equal within groups. To formalize this, consider a production component \( q_{n,l} \) on level \( l \). It is a function of components \( q_{m,l-1} \) associated with \( q_{n,l} \). If we isolate this production “branch”, we can interpret it as a production function on its own. Assuming prices \( p_{m,l-1} \) and \( p_{n,l} \) (corresponding to \( q_{m,l-1} \) and \( q_{n,l} \), respectively) exist, we might examine the problem

\[
\max \left[ p_{n,l}q_{n,l} - \sum_{m \in n} p_{n,l-1}q_{m,l-1} \right],
\]

subject to

\[ q_{n,l} = q_{n,l}(q_{m,l-1}; m \in n). \tag{5.66} \]

This formulation is completely analogous to the one in (5.3): it now refers only to a particular branch rather than to total production. In order to arrive at a total production function which exhibits constant returns to scale, we assume the same for every particular production branch; in mathematical terms

\[ \alpha q_{n,l} = q_{n,l}(\alpha q_{m,l-1}; m \in n), \]

for a finite (although possibly limited) range of \( \alpha \) around one. Problem (5.66) does not have a unique solution; we might, however, fix one quantity to determine the scale and proceed conditionally on this level of operation. A natural candidate for the scale quantity is the branch “output”, \( q_{n,l} \). In the same way as we define the elasticities of substitution for the overall production function, we can define the elasticities of substitution for the branchwise production function \( q_{n,l} \).

These “partial” elasticities of substitution tell us something about the substitution effects within the branch, holding branch output (represented by the component \( q_{n,l} \)) constant.

Now we assume that the production function \( f(\cdot) \) is a nested production function, where

(i) all branches exhibit constant returns to scale, at least locally, and
(ii) all partial cross-price elasticities of substitution are equal within a branch. We write \( \sigma_{n,l} \) for the partial cross-price elasticity of substitution of the branch \( q_{n,l} \).

These two assumptions imply that \( f(\cdot) \) might be locally approximated by the so-called nested-CES production function, where CES stands for constant elasticity of substitution. From the theory of the nested-CES production function, the following results are now directly available:\(^{11}\)

for \( n,m \neq s, \)

\[
\sigma_{nm} = \sigma_{n,K}c_{n,K}^{-1} - \sum_{l,K+1} \sigma_{n,l} [c_{n,l-1}^{-1} - c_{n,l}^{-1}], \quad n \neq m, \tag{5.67}
\]

\[
\sigma_{nn} = - \sum_{l} \sigma_{n,l} [c_{n,l-1}^{-1} - c_{n,l}^{-1}], \tag{5.68}
\]

where \( K \) equals the lowest level at which a component exists, associated with both \( q_{n} \) and \( q_{m} \) (the lowest common level) and \( c_{n,l} \) is defined by

\[
c_{n,l} = \sum_{m \subseteq n} c_{m}, \tag{5.69}
\]

\(^{11}\)See Moerland (1978) and Keller (1976).
i.e. the sum of the cost shares associated with the aggregate $q_{n,l}$ at level $l$ or, in other words, the cost share of component $q_{n,l}$.

Thus, in our example, if $q_n$ stands for the quantity of labour services, $c_{n,1}$ equals the cost share for value-added and $c_{n,2}$ the share for aggregate production, which is equal to one (so $c_{n,1} = -c_{n,2}$). If $q_m$ stands for capital services, the lowest common level for $q_n$ and $q_m$ equals one, since value-added is the lowest level aggregate including both capital and labour.

The interpretation of the elasticities of substitution, as given in eqs. (5.67) and (5.68), is similar to the interpretation in the household case [see eqs. (4.84) and (4.85)]. Instead of holding utility components constant, we now hold production components constant when considering the price effects per branch. Note that if, for a particular pair $n,m$, $L = K$ (the top level is the lowest common level), we have

$$\sigma_{nn} = \sigma_{n,1} = \sigma_{n,2}, \quad m \neq n; \quad m, n \neq s,$$

since $c_{n,1}$ equals one. If the production function has only one level of aggregation ($L = 1$), we have $K = L = 1$ for all $n$ and $m$ and, therefore,

$$\sigma_{nn} = \sigma_{n,1} = \sigma, \quad m \neq n; \quad m, n \neq s.$$

This case corresponds to the non-nested production function with elasticity of substitution $\sigma$, as considered in section 5 above.

Finally, the same remarks as made at the end of Chapter 4 also apply here, concerning the production function. Although the nesting enables us to reduce the number of unknown elasticities of substitution, we must specify a priori which elasticities are equal or, in other words, what the production structure looks like. This may be troublesome. On the other hand, our specification in terms of aggregates makes it possible to use estimates of parameters of aggregate systems, without having to resort to simultaneous estimation of the complete production structure.