Springer Texts in Business and Economics

For further volumes:
http://www.springer.com/series/10099

Svend Rasmussen

Production Economics

The Basic Theory of Production Optimisation

Second Edition

Springer

Svend Rasmussen
Department of Resource Economics and Food Policy
University of Copenhagen
Frederiksberg
Denmark

ISSN 2192-4333 ISSN 2192-4341 (electronic)
ISBN 978-3-642-30199-5 ISBN 978-3-642-30200-8 (eBook)
DOI 10.1007/978-3-642-30200-8
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012944979

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

In this second edition, Chap. 15 (Risk and Uncertainty) has been extended and includes now more on the theory of modelling risk and uncertainty. Further, a new Chap. 21 (Modelling Supply Functions Using Linear Programming) has been added, which includes a new example of using Linear Programming for production economic modelling. Finally, a number of misprints in the first edition have been corrected.

Copenhagen
April 2012

Svend Rasmussen

Preface to the first edition

This book has been written as a textbook for the course Production Economics and is as such aimed at students of economics and other students who are interested in studying production economic theory at the undergraduate level. It is recommended that the student has taken prior introductory courses in economics and has, therefore, obtained a sound initiation to economic thinking including graphic illustration and the analysis of (micro) economic issues. However, reading the book does not require knowledge of any specific economic theory.

The book adds to the existing literature in the sense that compared to the general microeconomic textbooks, which normally include a few chapters on production, cost, product supply, input demand and production under uncertainty, this book focuses on these subjects and treats them both graphically and mathematically in more detail. At the same time, it focuses on the application of the theory to solving illustrative problems related to production optimisation, and in this context it includes subjects which are normally not included in microeconomic textbooks like for instance optimisation of production over time and the use of linear programming for production optimisation.

Readers are encouraged to contact the author with any suggestions for potential improvement or regarding possible errors for the next edition, preferably by way of e-mail: sr@foi.ku.dk

Copenhagen
April 2010 Svend Rasmussen

Contents

1 Introduction

This book is concerned with production and related economic issues.

Generally speaking, *production* consists of the transformation of factors of production into products. The way in which the production is carried out – the production process – is outlined in Fig. 1.1.

Firstly, the *factors of production* (also called inputs) are taken to a *production plant*, which is where the actual production is carried out by way of a *production process,* the result of which being one or more *products* (also called outputs).

An example: Producing the product cereal involves adding the *factors of production* seeds, fertiliser, pesticides, labour, and machinery to the *production plant* land. The *production process* then takes place which includes the cultivation of the land, sowing, spraying, the waiting time required for the cereal crops to grow, and subsequent harvesting. The final result is two *products*: cereal (grain) and straw.

The economic issues related to production are based on the assumption that production takes place within the framework of what we call a *firm* (or company). The firm, in the classical sense, is an entity made up of production facilities (assets) owned by a physical person or a legal person (stockholder company), the *owner*, e*mployer* or *entrepreneur* who:

(a) Enters into a contract with each of the individuals who supply productive services. The contract specifies the nature and duration of these services and the remuneration required for them;
(b) Either makes decisions, or has the right to insist that decisions are made, in *her* interest, subject to her contractual obligations;
(c) Has the right to the *residual income* from production, i.e. the excess of revenue over payments to suppliers of productive services made under the terms of their contracts;
(d) Can transfer the right in the residual income, and her rights and obligations under the contracts with suppliers of productive services, to another individual;

S. Rasmussen, *Production Economics*, Springer Texts in Business and Economics,
DOI 10.1007/978-3-642-30200-8_1, © Springer-Verlag Berlin Heidelberg 2013

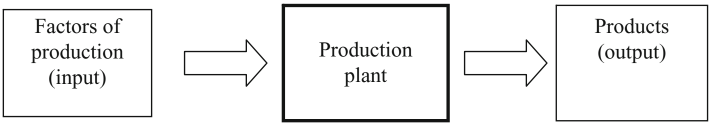


Fig. 1.1 Illustration of production

(e) Has the power to direct the activities of the suppliers of productive services, subject to the terms and conditions of their contracts;
(f) Can change the membership of the producing group not only by terminating contracts but also by entering into new contracts and adding to the group.
(Gravelle and Rees (2004), p. 93)

When talking about *the producer* in the following, it is this person (physical or legal) that is being referred to. It is the *decision maker*, who has the legal rights to the production facilities (because he owns or leases them), who is able to buy inputs and to decide what to produce, and who also carries the economic responsibility, in the sense that this person has the right to the residual income, i.e. the money remaining after all expenditures have been paid according to contracts with suppliers and other external parties.

The basic assumption is that it is the objective of *the producer* to maximise the gain (maximise the profit). The gain (the profit) is calculated as being the difference between the value of the produced products (the product value) and the value of the factors of production (costs) used. This objective is often called simply *profit maximisation*.

Based on the assumption of profit maximisation, three classical economic issues related to the act of producing can be identified:

1. *What to produce?* The producer usually has the option of producing alternative products with the available production plant. The farmer may grow e.g. barley or potatoes or oats on his/her land. He/she may either choose to grow all three crops, or choose to grow only one of them. However, what products would it be optimal to grow, i.e. what products would yield the highest profit?
2. *How much to produce*? A production process can be carried out more or less intensively. Crops can be grown using a larger or smaller amount of fertiliser, and when feeding livestock, a larger or smaller amount of fodder can be used. The size of the production will depend on this. But what is optimal? To add more fertiliser, which would result in a large production, or to add less fertiliser, which would result in reduced costs?
3. *How to produce*? A product can often be produced in several ways. When growing potatoes, for example, it is possible to fight weeds by the use of labour, herbicides or machinery. But what choice would be optimal? What kind of input would result in the lowest costs? *Time* is also an important factor. Should the farmer terminate fattening his slaughter pigs and send them to the slaughter house this week, or should he wait until next week?

When speaking of production and related economic issues it is often assumed that the production plant itself is given. If this was the case, the key economic issues concerning production would be related to the question of how to best utilise the given production plant. Should the gardener use the greenhouse to grow tomatoes or cucumbers? Should the farmer use his machinery and fixed family labour to grow potatoes or to produce Christmas trees?

However, in practice the economic issues concerning production are not that well-defined. In practice, it is of course possible to make *changes* to the given production plant, either by investing in new production facilities, or by renting (leasing) production facilities. A greenhouse can be viewed along the same lines as other factors of production, and the issue of how much "greenhouse" it would be optimal to apply, is in principal also an entirely ordinary production economic issue.

Whilst the answer is yes in principle, when it comes to decisions which have long term implications and concern the production *framework*, such issues are traditionally discussed within the discipline of investment and financial planning. This division is maintained in this book. However, there is no clear-cut distinction, and this book also includes theory for when the fixed asset and the related fixed costs become variable.

The description of the theory of optimisation of production is, in the majority of the book, based on the assumption that the price of inputs and outputs are determined by external factors and cannot be influenced by the producer. We say that the producer is a *price taker*. The book does, however, include a generalisation of the theory to account for conditions in which prices are not constant but dependent on the size of the production. Generally, there are no real problems in deriving principles for the optimisation under conditions in which prices are not fixed, i.e. they depend on the quantity produced. However, in this context, the problem of the pricing of output becomes an important subject. Problems relating to *pricing, marketing and the sale of products* are *not* discussed in this book. This comprehensive and for many companies important problem area, belongs to the subject area of market economics (industrial organisation). The reader is referred to other relevant textbooks to study this subject.

The theories and methods that are discussed in this book presuppose, in principle, complete certainty. It is important to be aware of the basic "building blocks" that a theory based on complete certainty entails before addressing the decision problems under risk and uncertainty. The subject of planning under risk and uncertainty is comprehensive and important, and the related methodological basis for this subject area is at present undergoing rapid development. A short introduction to the subject 'decision-making under risk and uncertainty' is given in Chap. 15, but students who want a more comprehensive treatment are referred to the extensive literature on this subject. It is relatively easy to derive criteria for optimal production decisions under uncertainty when the producer is assumed to be *risk neutral*. Under conditions in which the producer has *risk aversion,* it is difficult to derive useful criteria for the optimisation of production because it presupposes knowledge of the producer's preferences (utility function).

The content of the book is organised as follows:

Chapters 2, 3, 4, and 5 introduce the basic production economic tools. Chapter 2 begins with a description of the production function. Although Chap. 2 is purely technical and includes no discussion of behaviour/economics, it is the most important part of the book. The reason for this is that production economics, as presented in this book, is about how to maximise profit under the given circumstances. The *given circumstances* are the technical opportunities that the firm has, and the input and output prices it faces. Therefore, the three basic elements of production economics are *behaviour* (profit maximisation), *technology* (technical opportunities) and *prices* (input and output prices). The subject *behaviour* is not treated in this book, as we just assume profit maximising behaviour. *Prices* are assumed to be given from the outside (except in Chap. 13). This leaves *technology,* the form of which determines the economic results derived in the following chapters, and it is therefore important to carefully study the production function and its various forms.

Chapter 6 deals with the measurement of production. Although this is relevant only in a descriptive context, the subject is included here because it describes how to model changes in technology over time, and how to measure the production performance of firms. Therefore, this chapter is an important link to the descriptive approach to production economics, which is treated in more detail in other textbooks e.g. Chambers (1988).

Chapters 7, 8, and 9 show how it is possible to use the tools developed in the previous chapters to derive the firm's demand function for input, and its supply function for output. The chapter thus provides the microeconomic foundation for the analysis of demand and supply at the industry level.

Chapter 10 derives criteria for optimising production under restrictions and the mathematical tool used is the Lagrange function. The chapter describes how to use the concept of the pseudo scale line, introduced in Chap. 4, to analyse the adjustment of production when different types of production regulation are introduced. The chapter presents a number of different examples of production regulation and how the firm may adjust production in each case.

Chapter 11 introduces the concepts of economies of scale and economies of size. Whilst the two concepts are related, it is important to understand the difference between the two; economies of scale is a purely technical description of the production function, while economies of size is an economic concept which is useful for the discussion of the optimal firm size.

Chapter 12 returns to the concept of fixed production factors. It provides a formal definition of fixed production factors and describes why some production factors become fixed.

Chapter 13 relaxes the assumption of perfectly competitive markets, and it derives how firms facing a downward sloping demand curve (sales curve) for their products should optimise production. The chapter includes a description of perfect competition as the benchmark and the two cases of pure monopoly and monopolistic competition.

Chapter 14 gives an extensive introduction to the theory of how to optimise the production period. In this chapter, we introduce time as a new dimension of production, which may be dealt with by introducing time dated inputs and outputs to the previous models. However, to avoid the dimension problem in practical planning and to get operational solutions, it is often necessary to introduce simplifying assumptions when dealing with the optimisation of production over time.

Chapter 15 introduces risk and uncertainty and describes how to model these within a state-contingent framework. The expected utility model, which is a special case of the state-contingent model, is also described. Although risk and uncertainty is present in almost all production planning, it is often difficult to apply the theoretical models in practical planning because the decision maker's utility function is not known. Some of the ad hoc models used in practical planning under uncertainty are presented, but for a more thorough treatment of this subject the student is referred to other textbooks, for instance Rasmussen (2011).

Chapter 16 focuses on natural production factors such as agricultural land, and discusses the concept of economic rent. Although fixed factors of production provide the owner economic rent, the economic rent is often capitalised, meaning that producers who want to acquire some of these production factors from other producers pay a price, which passes on some, or all, of the economic rent to the seller, such that the net gain to the new owner is zero. The economic rent model is an important tool to describe the pricing of scarce resources such as land and production permits. Therefore, this model is also relevant when analysing the consequences of production regulation.

Chapter 17 can be considered as an introduction to the subsequent three chapters. It describes how producers should allocate fixed resources to various products when the producer has the opportunity to produce more products, or to produce products in different ways. The chapter generalises the theory of optimal use of input and output to the multi-input, multi-output case.

Chapters 18, 19, 20, and 21 introduce Linear Programming (LP) as a useful, operational tool for production planning. While the results in the preceding chapters have been presented in a general and sometimes abstract form, the Linear Programming model introduced in Chaps. 18 and 19, and demonstrated in Chaps. 20 and 21, is a powerful tool developed after the Second World War and used for practical planning in many contexts. Linear Programming is the linear version of the more general tool Mathematical Programming. Even though Linear Programming is based on linear functions, it is even able to handle non-linear cases, especially when combined with the facility of integer programming. To the applied production economist, the material in these three chapters is essential. In order to operationalize Linear Programming, one needs appropriate software. However, I have decided not to present any specific software in this book as there are so many, and so it is up to the individual to choose appropriate software for the task. There is a lot of good software available on the market, including the programme which probably everybody knows, Microsoft Excel. Specialised software such as LINDO is excellent for beginners, whilst more advanced users probably prefer GAMS.

The book includes an appendix on profit concepts. Although calculation of profit may seem straight forward, it is often not as simple as one may think. Students who have studied *accounting* will already know how to calculate profit in an accounting context. However, accounting is concerned with the past, whilst production economics is about *planning* for the future, and in this context the relevant cost concept is *opportunity costs*, which may be quite different from the costs registered in accounts.

I think that it is appropriate to conclude this chapter with reference to the following two rules, borrowed from Reekie and Crook (1995), which summarise the essence of this book:

1. A course of action should be pursued until its marginal benefits equal its marginal costs, that is, where marginal net benefits equal zero.
2. If no action can be pursued to the optimum extent, each different action should be pursued until they all yield the same marginal benefits per unit of cost.

References

Chambers, R. G. (1988). *Applied production analysis: A dual approach*. New York: Cambridge University Press.

Gravelle, H., & Rees, R. (2004). *Microeconomics* (3rd ed.). Harlow: Pearson education.

Rasmussen, S. (2011). *Optimisation of production under uncertainty. The state-contingent approach* (Springer Briefs in Economics). Heidelberg: Springer.

Reekie, W. D., & Crook, J. N. (1995). *Managerial economics* (4th ed.). Cambridge: Prentice Hall.

2 The Production Function

2.1 Introduction

Economic theory is, to a large extent, about money – about costs, prices, markets, return on investment, profit and similar economic concepts. This is also the case with the theory of production economics. However, the theory of production economics is special in that the limits of economic behaviour are defined by the *technical* production possibilities. Production technology is the decisive factor regarding the quantity produced and how it may be produced. Therefore, a very important part of the theory of production economics consists of describing the production technology which defines the framework for the economic behaviour.

This chapter is concerned with the description of production technology, which is traditionally based on the production function. Apart from the production function, the chapter also introduces a number of other concepts related to the description of production technology.

2.2 Production Technology

Production technology is, in its most general form, a description of the relationship between input and produced output. The description of production technical relationships is based on empirical observation of relationships between inputs and outputs, as e.g. described in Table 2.1 which shows the relationship between the addition of nitrogen fertiliser (N) and the cereal yield.

The specified relationship can be illustrated graphically, as shown in Fig. 2.1 next to the table. *It is this curve, as shown in Fig. 2.1 that was first referred to as a production function.*

Later, in line with the development of mathematical and statistical tools for the description of production technical and economic relationships, the production function was described by means of mathematical function relationships. The choice of the functional form to illustrate the empirically observed relationships

S. Rasmussen, *Production Economics*, Springer Texts in Business and Economics,
DOI 10.1007/978-3-642-30200-8_2,

Table 2.1 Yield with increased N addition

Kg of nitrogen (N) per hectare	Cereal yield, units per hectare
0	15
30	25
60	45
90	70
120	85
150	85
180	75

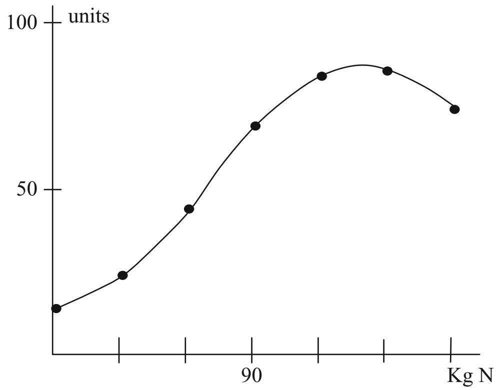


Fig. 2.1 Production function

as a nice curve which would pass through the observed points, as shown in Fig. 2.1, and the subsequent estimation of the parameters of the function itself, came to be an important discipline in production economics.

However, to describe production technology based *solely* on observations of relationships between inputs and outputs, as shown in Fig. 2.1, is inadequate.

Firstly, it should be noted that the curve in Fig. 2.1 only describes the quantity of produced output as the function of *one* input. However, what about the other inputs used in the production? Apart from nitrogen fertiliser, the use of labour, seeds, fuel, machinery etc. is also required when growing cereal crops. Generally, production always includes at least two, and often more, inputs. A complete description of the production technology for a given product will therefore presuppose a multi-dimensional illustration providing a *simultaneous illustration* of the relationship between output and *all* inputs. With a certain level of drawing skill, such a graphical illustration is possible for productions with only two inputs. However, this is not possible if there are three or more inputs. The solution could be to describe the production technology as partial production functions, i.e. functions with only one variable input, while the remaining ones are presumed to be fixed at a given level. With e.g. eight inputs, this would require that the production technology should be illustrated as eight figures similar to Fig. 2.1. Such an illustration is, however, insufficient since the interaction between the various inputs is unclear from these partial figures.

Secondly, it is not possible to be certain that the described relationships between inputs and outputs, as shown in Table 2.1, constitute a *complete* description of the production technology. Can one, for instance, be certain that there are no other ways to produce 45 units of cereal crops than by the exact application of 60 kg of nitrogen fertiliser? What if the observations in Table 2.1 originate from a producer, who is not technically efficient, i.e. produces less for a given input level than that which is technically possible? In such a case, there would be other possible points above the curve in Fig. 2.1, which should therefore also be included to give a complete description of the production technology. The same would be true for the points below the curve. For example, is it not technically possible to produce 45 units of cereal crops through the use of 90 kg of nitrogen? Thus, the points below the curve should also be included to give a complete description of the production technology.

This shows that the act of describing the production function solely as a curve interlinking empirical observations of relationships between inputs and outputs may be much too *incomplete* and too *imprecise* a description of the production technology for a given product. The correct approach must be to describe the production technology as the complete set of all the actual possibilities at the producer's disposal.

However, how can the complete set be described in a precise and unambiguous way? How can a production technology be described in a way which leaves all possibilities open to the producer to put his/her production together in a way that is optimal for the person in question? And furthermore, how can the production technology be described in a way that makes it possible to explain empirical observations which are outside the production function in Fig. 2.1?

The strictly general point of reference would be to describe the actual possible combinations of inputs and outputs. If this set is called T, then T can be defined as:

$$\mathrm{T}(x, y) \equiv \{(x, y) : x \text{ can produce } y\} \tag{2.1}$$

in which T is the *technology set*, x is the amount of input and y is the amount of output. In this strictly general formulation, both x and y could be scalars or vectors. However, for now, both x and y should be considered as scalars (one input and one output, as in Fig. 2.1).

Looking at the production as described in Table 2.1 and Fig. 2.1, it is evident that the points $(x, y) =$ (0, 15), (30, 25), (60, 45), (90, 70), (120, 85), (150, 85), and (180, 75) all belong to T as it has in fact been observed that, for these combinations of x and y, x can produce y. Furthermore, the individual points in Fig. 2.1 are connected as a smooth curve indicating that these intermediate points are also possible and therefore belong to T. By doing this, it is presumed that x can be applied in any amount (x is infinitely divisible) and that the actual observations between the already plotted points will be distributed on an even curve through the points.

However, are there other points in Fig. 2.1 that belong to T? Yes, if it is possible to produce 70 units of cereal crops with 90 kg of nitrogen (which it is according to Table 2.1), then it ought also to be possible to produce less – e.g. 45 units of cereal crops – with 90 kg of nitrogen. The reason is that under all circumstances it is possible to take the 90 kg of nitrogen and dispose of the 30 kg so that the amount actually added would be 60 kg. And with 60 kg it would of course be possible to produce 45 units of cereal crops, according to the table. A more realistic description would be to imagine an inefficient producer who, even with an addition of 90 kg of nitrogen, only achieves a yield of 45 units, exactly because the producer does not produce efficiently.

In a similar way it can be argued that *all the points below the curve* (but above the abscissa) in Fig. 2.1 also belong to T. The premise behind this argument is the possibility of *free disposability of input* or – which is a reference to the same – that there are producers who are not as efficient regarding their production as the most efficient producers on the actual production function.

What about the points above the curve? Do any of these belong to T? No, if the data used in Table 2.1 derives from an *efficient producer* there will be no possibility – with the technology under consideration – of achieving yields above the curve in Fig. 2.1. However, if the data used in Table 2.1 derives from a "poor" producer – a producer who, if he had been a little more meticulous with his production, would have produced a higher yield at each of the indicated input levels – then there would have been points above the curve in Fig. 2.1 belonging to T, as T includes the points where x can produce y. And if this is a matter of only having received data from a "poor" producer, and a "good" producer would have been able to achieve a higher yield, then there would in fact be points above the curve in Fig. 2.1 belonging to T.

The problem is not insignificant and may give rise to considerable problems and challenges in connection with production economic research that makes use of empirical data (data from the real world). As it is, such data come from producers who are different, some of whom are "good" while others are "poor". This being the case, the challenge is to establish which of these data do in fact make up the "border" of T (efficient producers) and which data derive from producers below the curve. The extent of the problem grows when the number of inputs (and outputs) increases to more than one.

This concludes the discussion of this issue. Notice that if the upper limit of T should be identical with the production function as illustrated in Fig. 2.1, then it presupposes that the data for the description of this function derives from an efficient producer.

2.3 The Input Set

The technology set T illustrates all the possible combinations of input and output. However, the production technology can also be defined in another way, i.e. as the *input set* $X(y)$ which is defined as:

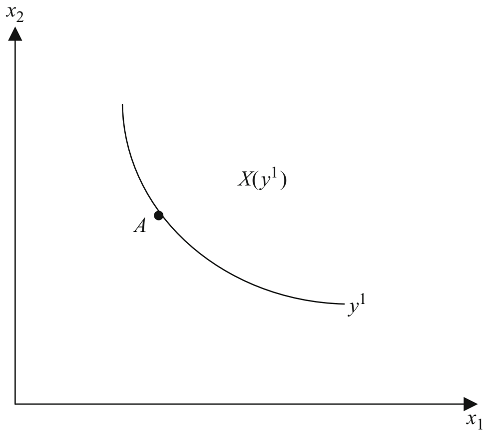


Fig. 2.2 Isoquant and input set

$$X(y) = \{x : x \, \text{can produce} \, y\} \tag{2.2}$$

The input set $X(y)$ attaches to each value of y the amounts of input x that can produce y. If Fig. 2.1 is used again with the choice of a y value, e.g. $y = 70$, it is obvious that the input amount of 90 kg N can produce 70 units of cereal crops, i.e. $90 \in X(70)$. However, if 90 kg N can produce 70 units of cereal crops then a larger amount of N can also produce 70 units of cereal crops when the precondition of free disposability of input is applied. Hence, the set of x's which can produce 70 units of cereal crops consists of those amounts where $x \geq 90$, i.e. $X(70) = \{x: x \geq 90\}$. If all values of y are considered it would result in an illustration of the same technology sets as described in T.

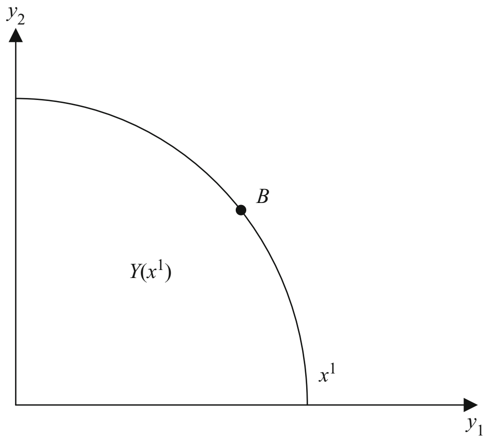
The input set can also be illustrated graphically when there are two inputs. Figure 2.2 shows an isoquant for production of the product y in the amount y^1 using the two inputs x_1 and x_2. An isoquant consists of those combinations of x_1 and x_2 that can produce the given product amount y^1. Hence, the point A illustrates an input combination which can in fact produce the amount y^1.

However, if such amounts of x_1 and x_2 – for instance corresponding to point A in Fig. 2.2 – can produce y^1 then larger amounts of x_1 and x_2 will also be able to produce the amount y^1 on the precondition of the existence of free disposability of input. Hence, the amounts that can produce y^1 ($X(y^1)$) are equal to the x's that are placed on and north-east of the isoquant in the figure.

2.4 The Output Set

The production technology can also be described by considering the product amounts which may be produced by a given input amount $x = x^1$. *The output set* $Y(x)$ is defined as:

Fig. 2.3 Production possibility curve and the output set



$$Y(x) = \{y : x \,\text{can produce}\, y\} \tag{2.3}$$

The output set $Y(x)$ attaches to each value of x the amount of outputs that can be produced by use of the given amount of inputs. In Fig. 2.1, the amount of outputs that can be produced using 60 kg of nitrogen equals 45 units of cereal crops, i.e. $45 \in Y(60)$ in any case. However, if it is possible to produce 45 units of cereal crops with 60 kg of nitrogen, then it is also possible to produce *smaller* amounts of output with 60 kg of nitrogen. It would under all circumstances still be possible to produce the 45 units of cereal crops and then subsequently dispose of a part of the produced amount! Hence, on the precondition of *free disposability of output*, the set of y's that can be produced with 60 kg of nitrogen consists of those amounts where $y \leq 45$, i.e. $Y(60) = \{y: y \leq 45\}$.

The output set can also be illustrated graphically when there are two outputs. Figure 2.3 shows a production possibility curve for the production of the two products y_1 and y_2 with a given input amount x^1. The production possibility curve consists of those combinations of y_1 and y_2 that can be produced with a given input amount x^1. Hence, point B illustrates the output combination that can be produced with the input amount x^1.

However, if the amounts of y_1 and y_2 corresponding to point B can be produced by x^1, then smaller amounts of y_1 and y_2 could also be produced by the input amount x^1 on the precondition of the existence of free disposability of output. Hence, the amounts that can be produced by x^1 ($Y(x^1)$) are equal to the y's that are placed on and south-west of the production possibility curve and limited by the coordinate system axes.

2.5 The Production Function

With the definitions of the technology set, input set and output set presented in the above section in place, it is now possible to give a more formal and precise definition of a production function than the definition associated with the "empirical" production function described in Fig. 2.1. The following definition presupposes that y is a scalar (an output), while x is a scalar or a vector of input:

Definition. A production function f is defined as:

$$f(x) = \max\{y : y \in Y(x)\} \tag{2.4}$$

The production function could also be defined as:

$$f(x) = \max\{y : y \in \mathrm{T}(x, y)\} \tag{2.5}$$

Hence, a production function is defined as the maximum amount of output that can be produced (through the use of a given production technology) with a given amount of input.

Similarly, isoquants and production possibility curves can be given formal definitions. An *isoquant is defined* as "the border" of the input set, i.e. as the x's for which the following is true:

$$G(y) = \{x : x \in X(y) \mid x^k \notin X(y) \text{ for } x^k \leqslant x\} \tag{2.6}$$

in which $x^k \leq x$ is to be understood as: None of the elements (x_i) in the vector x^k are greater than the corresponding elements in the vector x, and at least one of the elements in x^k is smaller than the similar element in x.

If the possibility of production of multiple outputs exists, then the production possibility curve is defined similarly as:

$$P(x) = \{y : y \in Y(x) \mid y^k \notin Y(x) \text{ for } y^k \geqslant y\} \tag{2.7}$$

in which $y^k \geq y$ is to be understood as: None of the elements (y_i) in the vector y^k are smaller than the corresponding elements in the vector y, and at least one of the elements in y^k is greater than the similar element in y.

2.6 Diminishing Marginal Returns

Following this strictly formal definition of the production technology and production function, we shall now return to the graphical illustration of the production function which was the point of reference in Fig. 2.1 at the beginning of the chapter. But what would a purely graphical version of the production function look like? And what about the mathematical representation of the production function? What kinds of functions are used to represent production functions?

2.6.1 The Law of Diminishing Marginal Returns

First we will have a look at the graphical representation of a production function.

Recall that a production function can only be drawn on a piece of paper if there is one or at the most two inputs. As more than two inputs are normally used in a production, (almost) all graphical illustrations of production functions presuppose the presence of one or more underlying inputs (part of the production) with given fixed amounts (fixed input). The curve illustrating the relationship between added nitrogen fertiliser and the yield of cereal crops in Fig. 2.1, therefore, presupposes that all the other inputs used in the production of cereal crops (seeds, pesticides, land, labour, machinery, etc.) are present in given fixed amounts.

An essential precondition related to a production function is the assumption of *diminishing marginal returns*. The precondition, which is based on empirical observations of how the production is carried out in practice, is universally acknowledged as a basic condition within production economics referred to as the *Law of diminishing marginal returns*. Briefly explained,

> *The Law of diminishing marginal returns* states that by adding increasing amounts of input to a production with at least one fixed input, the additional returns resulting from the addition of increasing amounts of input will gradually diminish, and eventually become negative.

The concept of marginal returns is used here to refer to the increase in production arising from the addition of an extra unit of input. Normally, this increase is expressed by the slope of the production function, i.e. as the value of the derivative, i.e. $df(x)/dx$, if x is a scalar, or the partial derivative, $\partial f(x)/\partial x_i$, if x is a vector. Expressed this way, the concept of *marginal returns* or *marginal product* is normally used to express the additional returns per input unit in connection with *marginal* changes in the amount of input.

If the function expression of the production function is unknown, the marginal product can be approximated by the use of the *difference product* expressed as $\Delta y/\Delta x$. Using data from the example in Table 2.1, the difference product in the interval from 30 to 60 kg of nitrogen equals (45–25)/(60–30) = 0.67, and in the interval from 90 to 120 kg of nitrogen equals (85–70)/(120–90) = 0.50. These difference products are approximated expressions of the derivative (and thereby the marginal product) at the centre of the relevant intervals.

The Law of diminishing marginal returns is nicely illustrated in the production function shown in Fig. 2.1. When adding small amounts of nitrogen fertiliser, the marginal product increases (the slope of the production function increases). At some point, the marginal product is diminishing, and when adding approximately 135 kg of nitrogen, the marginal product becomes zero and subsequently becomes negative with further additions. In this example, the precondition of at least one fixed input is satisfied as land and other inputs used in the production of cereal crops are presupposed to be present in given fixed amounts.

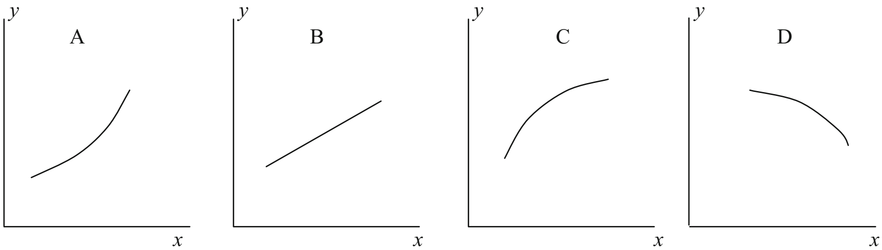


Fig. 2.4 Alternative production function shapes

2.6.2 Graphical Illustration of the Production Function

When production functions are represented graphically (and it is thereby presupposed that a number of underlying production factors are fixed inputs), such a representation will look the same as, or similar to, the curve in Fig. 2.1. These "similar" representations are produced by observing only parts of the shape of the total production function in Fig. 2.1.

Figure 2.4 illustrates four different (sub) shapes of the production function which are all contained in the production function outlined in Fig. 2.1. Example A outlines the progressively increasing shape with positive and increasing marginal returns. This shape corresponds to the first part of the production function in Fig. 2.1. Example B outlines a linear shape with positive and constant marginal returns, corresponding to the area between 60 and 90 kg N in Fig. 2.1. Example C outlines a digressively increasing shape with positive and diminishing marginal returns corresponding to the area between 90 and 130 kg N in Fig. 2.1. Finally, example D outlines a progressively diminishing shape with negative and diminishing marginal returns. This shape corresponds to the last part of the production function in Fig. 2.1.

A production function with all four "shape" types in the described order, like the one in Fig. 2.1, is referred to as the *neoclassical production function*. This type of production function has especially been used to describe production relationships within agriculture.

If you are not interested in the *overall shape* of the production function, but solely in the *local areas* of the production function, it is sufficient to plot the part of the production function that is of interest. As mentioned later on, the part of the production function that is of special interest in connection with production economics is the one that is illustrated in example C in Fig. 2.4 (the digressively increasing shape). Therefore, production functions are often illustrated graphically with a shape similar to example C in Fig. 2.4. However, this does not necessarily mean that this shape is present throughout the entire domain of the production function, i.e. *globally*. It might also solely be an issue of a description of a local shape.

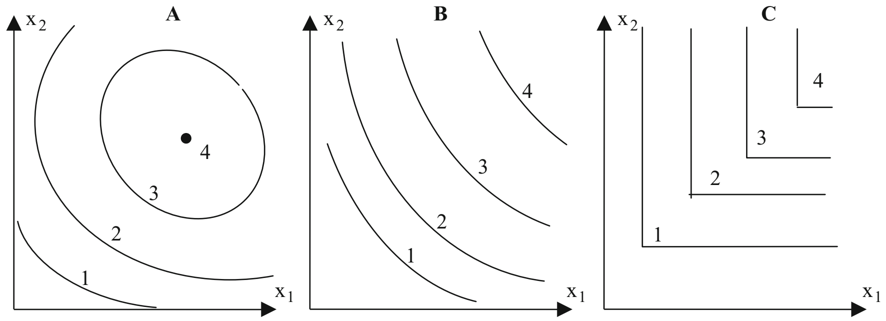


Fig. 2.5 Alternative sets of isoquants

If we consider production with more than one input, the graphical illustration of the production function is a bit more complicated. If there are two variable inputs, the production function is often described by means of so-called *isoquants* which are defined formally (for any number of inputs) in Eq. 2.6 and illustrated graphically as in Fig. 2.2 for two inputs. Isoquants can be interpreted as level curves for the production function. As the issue of interest regarding production economics is normally solely the area of the production function which corresponds to example C in Fig. 2.4, the similar areas of the isoquant will in fact consist of diminishing, convex curves, as illustrated in Fig. 2.2 (it is up to the reader to demonstrate why).

Figure 2.5 shows alternative sets of four isoquants. The number on each of the isoquants expresses the size of the production. In set A, the amount 1 can be produced with either input x_1 or with input x_2 or with a combination of x_1 and x_2. Hence, none of the inputs are *necessary inputs*. To produce amounts of 2, 3, or 4, both inputs are however necessary. The production function has a maximum of 4. In set B, both inputs are necessary and the production function does not have a maximum (this could e.g. be a Cobb-Douglas production function (discussed later)). Set C shows L-shaped isoquants on which only the corner points are efficient. This is a so-called Leontief production function (discussed later).

2.6.3 Mathematical Representation of the Production Function

The formal mathematical representation of the production function for the production of *one* output has previously been shown as in Eq. 2.4. Alternatively, Eq. 2.4 could be written as:

$$y = f(x) \tag{2.8}$$

in which y is a scalar (the amount of the product y), f is the production function, and x is a vector of inputs.

The production function:

$$y = f(x_1) \tag{2.9}$$

expresses the production of y only as a function of the variable input x_1. If it is appropriate to explicitly express that the production of output y is a function of the variable input x_1 and the fixed inputs $x_2, \ldots, x_n$, then the function (2.9) should be written as $y = f(x_1 \mid x_2, \ldots, x_n)$. Normally, fixed inputs are not included when writing the production function. It is however important to keep in mind that the production may depend on considerably more inputs than specified in the actual production function. Write y = f(x_1, x_2) or $y = f(x_1, x_2, \mid x_3, \ldots, x_n)$ if you want to express that the production is a function of two variable inputs.

There is no given mathematical functional form for a production function. All the functional forms that have been used to describe the production have historically been based on more or less subjective choices. The best known of these function forms is the so-called Cobb-Douglas production function which, with two variable inputs, has the form:

$$y = Ax_1^{b_1}x_2^{b_2} \tag{2.10}$$

in which A, b_1, and b_2 are predetermined parameters (constants).

Evidently, the choice of functional form depends on the areas of the production function which are to be described. Is it a global description which should cover the entire function shape as outlined in Fig. 2.1, or is it a matter of functions which should only illustrate local areas of the production shape, as e.g. illustrated by the four examples in Fig. 2.4? Hence, the Cobb-Douglas function is only capable of illustrating shapes such as the one shown in example C in Fig. 2.4. An alternative functional form, which also seems to be able to work here, is the simple quadratic function. In case of a linear shape as shown in example B in Fig. 2.4, it is possible to choose a simple linear function as the functional form.

The choice of functional form and the subsequent estimation of the parameters of the function is a comprehensive science in itself, which is not discussed in any further detail here. Anyone with a particular interest in this is referred to studies within the subject area of Econometrics.

2.6.4 The Production Elasticity

Apart from describing the production technology as a table with numerical relationships between inputs and outputs (Table 2.1), as a graph illustrating these numerical relationships (Fig. 2.1), or mathematically as an actual production function (Eqs. 2.9 and 2.10), it is possible to express these relationships between inputs and outputs locally by means of the so-called *production elasticity*.

The production elasticity expresses the relative change in production through a relative change in the addition of input. If e.g. 5 % more input is added and 4 % more output is achieved, then the production elasticity is 4/5 = 0.80.

If there are multiple inputs, it is possible to calculate the production elasticity for each input. Formally, the production elasticity ε_i for input i is calculated as:

$$\varepsilon_i \equiv \frac{\frac{\partial f(\mathbf{x})}{f(\mathbf{x})}}{\frac{\partial x_i}{x_i}} = \frac{\frac{\partial f(\mathbf{x})}{\partial x_i}}{\frac{f(\mathbf{x})}{x_i}} = \frac{MPP_i}{APP_i} \tag{2.11}$$

in which *MPP* and *APP* represent the marginal product (*M*arginal *P*hysical *P*roduct) and the average product (*A*verage *P*hysical *P*roduct), respectively.

If the function expression for the production function is not known, then the production elasticity can be approximated by replacing marginal change (∂) in Eq. 2.11 by small, numerical change (Δ). Hence, an approximated expression for the production elasticity in the centre of the interval is achieved by calculating the following:

$$\varepsilon_i \cong \frac{\frac{\Delta y}{y}}{\frac{\Delta x_i}{x_i}}$$

If the data from the example in Table 2.1 is used, the production elasticity in the interval 30–60 kg N is approximated using the calculation ε_i = [(45–25)/25]/[(60–30)/30] = 0.80. As the centre of the interval 30–60 kg is 45 kg, this elasticity (0.80) will be used as the approximated elasticity at the point where the 45 kg N are applied. Similarly, the elasticity at the point where the 105 kg N are applied is expressed as ε_i = [(85–70)/70]/[(120–90)/90] = 0.64. As shown by this example, the production elasticity (normally) depends on the point of reference, and the elasticity declines with the addition of input.

Some production functions have constant production elasticities. This is the case for the Cobb-Douglas production function shown in Eq. 2.10 in which the production elasticity for the input i (i = 1,2) is b_i (the reader is encouraged to verify this himself/herself using the expression after the second equal sign in Eq. 2.11 for the calculation).

3 Optimisation with One Input

3.1 Introduction

This chapter discusses the optimisation of production under the simplest preconditions: The production of one product (output) using one input. The amount of the other inputs is presumed given as fixed amounts. The prices of inputs and outputs are presumed given externally (the producer is a price taker) and these prices are presumed to be constant, no matter how much the producer buys and sells.[1]

The optimisation of the production takes place in two ways: Either by deciding how much input it is optimal to add, or by deciding how much output it is optimal to produce. The result (optimal values of x and y) is of course the same and the choice of one or the other method is a matter of preference.

The relationship between input and output is shown as a neoclassical production function in the upper half of Fig. 3.1. The lower half of Fig. 3.1 shows the corresponding curves representing the marginal product (*MPP*) and the average product (*APP*), respectively, as the function of the addition of the input x. The marginal product is equal to the slope of the production function and is formally defined as:

$$MPP = \mathrm{d}f(x)/dx,$$

while the average physical product equals the slope of a straight line through the zero point up to the production function and is formally defined as:

$$APP = f(x)/x = y/x.$$

[1] We refer to this as a *competitive market*. The non-competitive market case is discussed at the end of Chap. 7 (input) and in Chap. 13 (output).

S. Rasmussen, *Production Economics*, Springer Texts in Business and Economics,
DOI 10.1007/978-3-642-30200-8_3, © Springer-Verlag Berlin Heidelberg 2013

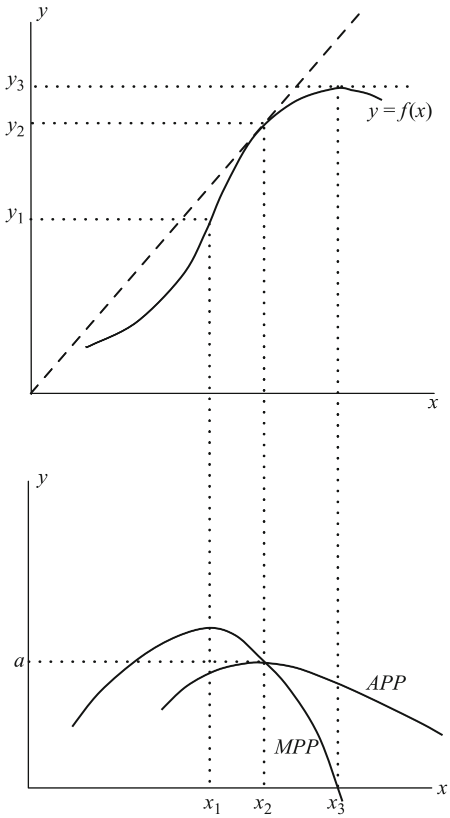


Fig. 3.1 The production function and *MPP* and *APP*

Based on this, we will first look at the optimisation of production from the input side.

3.2 Optimisation from the Input Side

When the optimal supply of input x is to be determined, it must initially be noticed that it will never be profitable (regardless of the input and output prices) to add larger amounts of input than the amount indicated by x_3 in Fig. 3.1. Larger amounts would result in decreasing output which would never be profitable with positive input and output prices.

Would it be possible in a similar way to determine a certain minimum amount of input x that should always (regardless of the input and output prices) be applied? – Yes, it would indeed. When the prices, as here, are presumed to be constant, it will – if it is at all profitable to produce the product in question – be optimal to apply an input amount that, as a minimum, corresponds to x_2 in Fig. 3.1.

To see why, presume that the price p of y is used as monetary unit so that the product price p equals 1. In this case, the production function in Fig. 3.1 is at the same time the measure of the *total product value* or *total revenue TR* ($TR = py = 1y = y$). Presume furthermore – as the point of reference – that the price of input (w) is such that the total costs of buying input (*total factor costs (TFC)*) ($TFC = wx$) are given by the dotted line, which is in fact tangent to the production function in Fig. 3.1. Under such circumstances, the use of the input amount x_2 would indeed result in a profit of 0 (nil) (total product value (TR) minus the total factor costs (TFC) equal to zero at y_2).

Now presume that the input price w is somewhat higher, so that the total factor cost follows a line with a larger slope than the dotted line in Fig. 3.1. If this is the case, it will not be profitable to produce anything at all as there will be no input amounts for which there is a positive profit ($TR\text{-}TFC < 0$).

Presume, on the other hand, that the input price is somewhat lower so that the total factor costs follow a line with a smaller slope than the dotted line in Fig. 3.1. and therefore intersects the production function (in two places). If this is the case, it will be profitable to produce as there are input amounts around x_2 where there is a positive profit ($TR\text{-}TFC > 0$). The input amount with the highest profit (largest distance between the production function and the line showing the total factor costs) is found in the area *to the right of* x_2.

Hence, it has been shown that with constant output prices, the optimal input supply is always to be found in the area of the production function corresponding to input amounts of between x_2 and x_3 in Fig. 3.1. It is with this observation as the basis that analyses of production economic issues are almost always limited to observing the part of the production function which corresponds to example C in Fig. 2.4.

After these introductory descriptions it is possible to analyse how the optimal input amount of input is formally determined.

The profit equals the difference between the total product value (or total revenue, TR) and the total factor costs (TFC). Hence, if the profit is referred to as π the profit is:

$$\pi = TR - TFC = py - wx = pf(x) - wx \tag{3.1}$$

in which p is the output price and w is the input price.

The maximum of π with respect to x is found when the derivative of π with respect to x is zero. If the right hand side of Eq. 3.1 is differentiated with respect to x and set equal to zero, this will result in the following equation for the determination of the optimal x:

$$p(\mathrm{d}f(x)/\mathrm{d}x) \equiv pMPP \equiv VMP = w \tag{3.2}$$

in which VMP is the value of the marginal product defined as the marginal product MPP multiplied by the product price p.

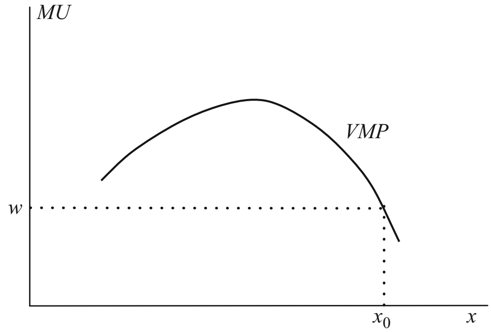


Fig. 3.2 Optimal input supply

The condition (3.2) states that to achieve an optimal input supply x, the value of the marginal product *VMP* (the increased value of the production at the marginal supply of one more input unit) must be equal to the input price w. This ratio is illustrated graphically in Fig. 3.2 in which x_0 is the optimal input supply.

The procedure is illustrated by the following example:

Example 3.1

The production function, $y = f(x) = 16 + 0.8x - 0.005x^2$

The product price, $p = 90$. The input price, $w = 5$.

The marginal product, $MPP = \mathrm{d}f(x)/\mathrm{d}x = 0.8 - 0.01x$

The value of the marginal product, $VMP = MPP \times p = (0.8 - 0.01x) \times 90 = 72 - 0.90x$

Optimal application of x when $VMP = w$, i.e. when: $72 - 0.90x = 5$, i.e. when $x = 74.44$

Optimal production of $y = 16 + 0.8 \times 74.44 - 0.005 \times 74.44^2 = 47.84$.

By re-writing Eq. 3.2 the optimal condition can also be written as:

$$MPP = w/p \tag{3.3}$$

As *MPP* is the slope of the production function, the optimal input supply x may therefore be found by drawing a line with the slope w/p and by letting this line be tangent to the production function. The input amount at this tangent point equals the optimal application of x.

The condition of optimality illustrated by the form Eq. 3.3 is expedient for use in a graphical analysis of what happens with the optimal addition of one input when the relations between prices change. It appears directly from Eq. 3.3 that if the input price increases compared to the output price, then the optimal input amount is found at a lower input amount as the slope of the production function is steeper here. This is illustrated in Fig. 3.3 in which the slope (the ratio w/p) in point A is relatively low, which results in an optimal input supply corresponding to the amount a. At point B, the slope is higher, which results in a lower optimal supply (b) of input.

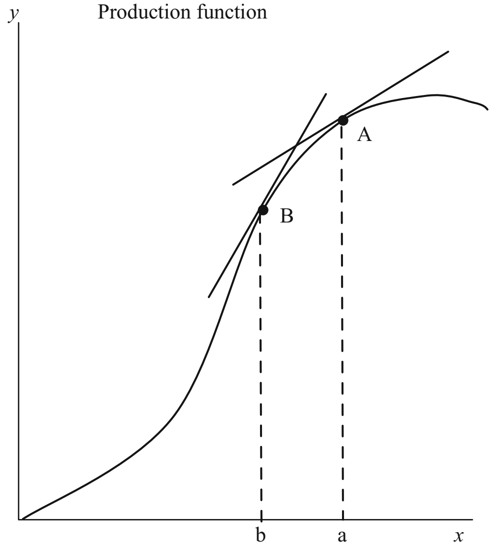


Fig. 3.3 Optimisation of input

The condition (3.2) stipulates only the *necessary* condition (the first order condition) for optimal input. The *sufficient* condition is found by adding the second order condition, as the maximum of a function presupposes that the *second derivative is negative*. If Eq. 3.1 is differentiated two times with regard to x and the condition is formulated so that the second derivative must be negative, this will generate the condition (3.4) which, together with Eq. 3.2, results in the sufficient condition for optimal input supply.

$$\mathrm{d}MPP/\mathrm{d}x < 0 \tag{3.4}$$

According to Eq. 3.4, the optimal input supply is therefore to be found for values of x when the marginal product is diminishing, i.e. for values of x which are higher than x_1 in Fig. 3.1.

When it comes to optimisation in practice, the mathematical form for the production function is often unknown. The production function exists solely as a table showing relationships between discrete values of input (x) and output (y) (e.g. similar to Table 2.1). If this is the case, the derivative, and therefore the marginal product (MPP), cannot be derived and hence it is not possible to use the condition of optimality 3.2.

Under such circumstances, *the difference product* is used as an approximated expression for the marginal product. The difference product is calculated as $\Delta y/\Delta x$, where Δx expresses the change in x (the difference between two "adjacent values" of x) and Δy expresses the corresponding change of y (the difference between the corresponding "adjacent values" of y).

Example 3.2

The following empirical example illustrates a situation in which the underlying production function corresponds to the production function in Example 3.1, but where only the discrete numbers in the first two columns are given. The exact marginal product (calculated as shown in Example 3.1) is included in the third column. The fourth column shows the difference product which is an approximated expression of the marginal product at the centre of the interval (hence, the number 0.55 is an estimate of the marginal product when x equals 25) etc.

		MPP (exact)	*MPP* (approximated)	*VMP* (approximated)
x	y $(=f(x))$	$\mathrm{d}f(x)/\mathrm{d}x$	$\Delta y/\Delta x$	$MPP \times p$
10	23.5	0.7		
			0.65	58.5
20	30.0	0.6		
			0.55	49.5
30	35.5	0.5		
			0.45	40.5
40	40.0	0.4		
			0.35	31.5
50	43.5	0.3		
			0.25	22.5
60	46.0	0.2		
			0.15	13.5
70	47.5	0.1		
			0.05	4.5
80	48.0	0		

With the information available, it is not possible to identify the exact optimum ($x = 74.44$ as shown in Example 3.1). The closest approximation achievable is that the optimal solution is to be found within the interval $70 < x < 80$, because it is in this interval that *VMP* (4.5) is closest to the input price w (5). The production in this interval is $47.5 < y < 48.0$.

3.3 Optimisation from the Output Side

When optimising the production as seen from the output side, the equation for the profit π is formulated as a function of y, and *not* as a function of x as in Eq. 3.1.

$$\pi = TR - TFC = py - wx = py - wf^{-1}(y) \tag{3.5}$$

in which f^{-1} is the inverse production function.

The inverse to a function only exists if the function is monotonous, i.e. either increasing *or* decreasing. The production function f in Fig. 3.1 has an increasing, as well as a decreasing shape, which is why the condition for the existence of a unique inverse function is not fulfilled. However, as illustrated in Sect. 3.2, the optimal production is always to be found for input amounts smaller than less x_3 in Fig. 3.1, i.e. on the increasing part of the production function. It will, therefore, be sufficient to observe the production function for values of x less than x_3. And in this area, the inverse to the production function is uniquely defined.

The expression $wf^{-1}(y)$ can be expressed more generally as the function c, such that:

$$c(w, y) = wf^{-1}(y) \tag{3.6}$$

The function c in Eq. 3.6 is referred to as *the cost function*. Basically, a cost function is defined as a function expressing the *lowest* costs by which the product amount y can be produced when the input price is w. As illustrated in the following sections, this definition is also true when it comes to multiple inputs and outputs, i.e. when y and w are vectors, and not just scalars.

Is it possible to be certain that c, as expressed in Eq. 3.6, does in fact express the *lowest* costs of production of y? Yes, it is. The production function f is in fact defined as a function which produces the *maximum* of y for each value of x. Therefore, the inverse function expresses the smallest amount of input whereby the product amount y can be produced.

The profit π in Eq. 3.5 can now be expressed as:

$$\pi = TR - VC = py - c(w, y) \tag{3.7}$$

in which VC expresses the variable costs involved in the production of y.

The maximum of π with regard to y is found when the derivative of π with regard to y is zero. If the right hand side of Eq. 3.7 is differentiated with regard to y and set equal to zero it will result in the following equation for the determination of the optimal y:

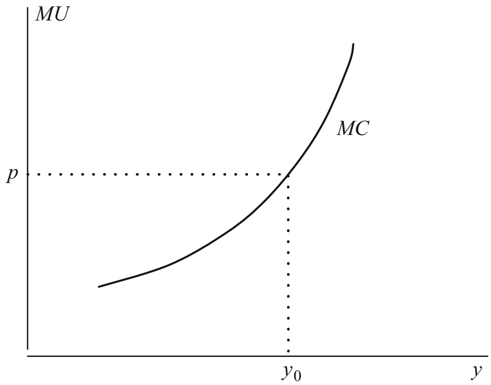
$$p = \mathrm{d}c(w, y)/\mathrm{d}y \equiv MC \tag{3.8}$$

in which MC is the abbreviation for the marginal costs, i.e. the incremental costs for the production of one additional unit of y.

Hence, the condition (3.8) states that optimal production is characterised by the product price p being equal to the marginal costs MC. The condition of optimality is outlined graphically in Fig. 3.4, in which y_o is the optimal production.

The shape of the marginal costs curve as a progressively increasing function of y is strictly related to the shape of the production function. In Chap. 5, we will return to this issue again through a thorough analysis of the cost function. For now, it should merely be established that, concerning the relevant area of the production function (C in Fig. 2.4), the shape of the marginal costs curve is progressively increasing as illustrated in Fig. 3.4.

Fig. 3.4 Optimal production of *y*



For the sake of completeness, it should be noted that the optimal production as calculated in Fig. 3.4 corresponds to the production achieved by the use of the optimal input amount as determined from the input side in Fig. 3.2.

Example 3.3

We use the same numerical example as in Example 3.2. This point of reference is now the production y and the corresponding (variable) costs c. The cost c is the result of multiplying the input price w and the applied input amount x. If the applied input amount is presumed to be the lowest amount of input x for the production of the relevant amount of y, then the lowest input amount is a unique function of y when production is carried out within the rational production area ($x < x_3$ in Fig. 3.1). This function is referred to as $x^*(y)$, where the asterisk (*) refers to the use of an optimised expression. With the assumption mentioned, the cost c can be expressed as $c = w \times x^*(y)$.

In this example, the marginal cost cannot be calculated directly, as the functional form of *c(w,y)* is not known (it could, in principle, be derived based on the inverse of the production function (see the Eq. 3.6), if desired). Therefore, the marginal cost must be approximated by the calculation of differences. *The difference cost* equals $\Delta c/\Delta y$, i.e. the change in costs divided by the corresponding change in production. The marginal costs (*MC*) approximated in this way are shown in the right hand column of the table below.

	Costs	*MC* (approximated)
y	$c\ (w \times x^*(y))$	$\Delta c/\Delta y$
23.5	50	
		7.69
30.0	100	
		9.09
35.5	150	
		11.11

(continued)

	Costs	*MC* (approximated)
y	$c\ (w \times x^{*}(y))$	$\Delta c/\Delta y$
40.0	200	
		14.29
43.5	250	
		20.00
46.0	300	
		33.33
47.5	350	
		100.00
48.0	400	

The information presented is insufficient to identify the optimum ($y = 47.84$ as shown in Example 3.1). The closest approximation achievable is that the optimal solution is to be found within the interval $47.5 < y < 48.0$, in that this is where *MC* (100.00) is closest to the output price (90). The cost in this interval is $350 < c < 400$.

As illustrated by a comparison of Examples 3.2 and 3.3, optimisation from the input side and from the output side generates the same result.

The relationship between the optimisation from the input side and from the output side is relatively simple in cases with only one input and one output. The relationship becomes more complicated as soon as multiple inputs or outputs are introduced.

With multiple inputs (and one output), one can think of two possible cases:

1. One possibility is that a cost function is known. Either in the form of an actual function expression for $c(w, y)$, or in the form of a table with numerical relationships between the output y and the costs c, as in Example 3.3. Here optimisation is carried out by identifying the value of y, where the marginal costs *MC* equal the output price p (optimisation as in Example 3.3)
2. The other possibility is that a production function, $y = f(x_1, x_2, \ldots)$ and the input $(w_1, w_2, \ldots)$ and output (p) prices have been given. *The optimisation is now a two step procedure: Firstly, it is decided – for each possible value of y – how the selected output amount y is produced with the lowest costs. This corresponds to determining the cost function. Then production is optimised as under item 1 by finding the value for y when MC = p.*

In Chap. 4 below, the first of the two steps mentioned under item 2 is discussed. Next, in Chap. 5, the second of the two steps mentioned under item 2 is discussed. Hence, Chap. 5 is a continuation of Sect. 3.3 of this chapter.

Production and Optimisation with Two or More Inputs

4

4.1 Introduction

In the real world, no production is carried out using only one input. Normally, several (controllable) inputs are used. Hence, when growing cereal crops, land, seeds, labour, fertiliser, pesticides, machinery, etc. are used. A car manufacturer uses steel, labour, leather, plastic, paint, tyres, fuel, etc. Various inputs can often replace each other so that it is possible to replace some of the expensive ones with cheaper alternatives if the price of one input increases. For example, if the price of pesticides, which are used to chemically control weeds in the field, rises, then the use of labour might be considered as an alternative to control the weeds. If the price of fuel used for heating factory or office buildings increases, it may be cheaper to use electricity for heating instead. The question as to the extent to which the various inputs can replace each other becomes the key question in this connection. This chapter deals with the instruments which can be used to address such issues. As in Chap. 3, we assume competitive input and output markets.[1]

The chapter primarily discusses issues concerning production optimisation with two (variable) inputs. Results concerning two variable inputs can easily be generalised to cover more inputs, and such a generalisation will be undertaken as part of this chapter.

When one of the two (variable) inputs becomes a fixed input, special conditions apply. The discussion of such conditions will, among other things, be relevant when issues concerning production adjustment under *restrictions* are discussed later on. Issues concerning production optimisation with input quotas are addressed in Chap. 10. The basic foundation for such analyses of production adjustment under production regulation is presented in this chapter.

[1] The non-competitive case is treated in Chap. 13.

S. Rasmussen, *Production Economics*, Springer Texts in Business and Economics,
DOI 10.1007/978-3-642-30200-8_4,

4.2 Cost Minimisation

The underlying basis for the following analysis is the production function $y = f(x_1, x_2)$ as illustrated by a set of *isoquants*. Figure 4.1 shows such a set of isoquants with three yield levels y^1, y^2, and y^3, when $y^1 < y^2 < y^3$.

The expansion path is defined as the curve connecting the points of the isoquants with the slope $-w_1/w_2$, where w_1 and w_2 are the prices of input 1 and input 2, respectively.

The expansion path is found by addressing the following formal problem:

$$\begin{array}{l}\min\{w_1x_1 + w_2x_2\} \\ x_1, x_2\end{array} \tag{4.1a}$$

under the constraint that:

$$y = f(x_1, x_2) \tag{4.1b}$$

The problem Eqs. 4.1a and 4.1b consists of minimising the variable costs Eq. 4.1a under the constraint that the amount y (4.1b) is being produced. The solution is found by using the Lagrange method (see Chiang (1984) p. 372). Firstly, the Lagrange function L is created:

$$L = w_1x_1 + w_2x_2 + \lambda(y - f(x_1, x_2)) \tag{4.2}$$

and minimised with regard to the two variables x_1 and x_2 and the Lagrange multiplier λ by taking the partial derivatives and setting them equal to zero. This will produce the following three conditions for an optimal solution to Eqs. 4.1a and 4.1b:

$$w_1 = \lambda \cdot MPP_1 \tag{4.3a}$$

$$w_2 = \lambda \cdot MPP_2 \tag{4.3b}$$

$$y = f(x_1, x_2) \tag{4.3c}$$

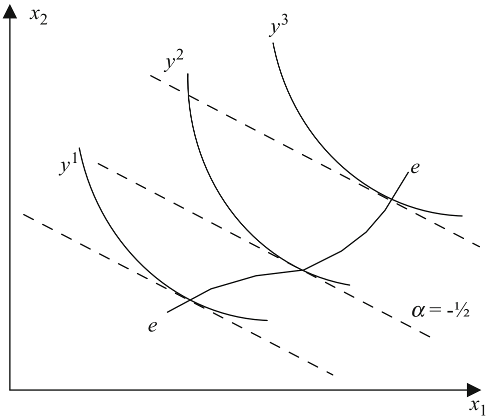
Dividing Eq. 4.3a by Eq. 4.3b produces the *necessary condition* for the minimisation of Eq. 4.1a for the given y:

$$\frac{w_1}{w_2} = \frac{MPP_1}{MPP_2} \tag{4.4}$$

This condition can be interpreted graphically as the tangent point between the so-called *budget line* and the isoquant for the production y. To find out why, consider the variable costs:

$$C = w_1x_1 + w_2x_2 \tag{4.5}$$

Fig. 4.1 Isoquants and expansion path



which should be minimised for the given production y according to Eq. 4.1a.

Finding the solution to Eq. 4.5 for x_2 produces:

$$x_2 = -\frac{w_1}{w_2}x_1 + \frac{C}{w_2} \tag{4.6}$$

which is a straight line in the $x_1 - x_2$ plane with the slope $-\frac{w_1}{w_2}$ and intersection point with the x_2 axis corresponding to $\frac{C}{w_2}$. This type of line is called *the isocost line*, as it presents combinations of x_1 and x_2 which all have the same costs, C. The concept, *budget line,* can also be used as this line presents a *budget constraint* in cases where there is only a limited amount of money (budget) C available to buy input. (You will learn more about the use of the isocost line and its interpretation in the section on optimisation under constraints in Chap. 10).

If the isoquant for a given production $y = f(x_1,x_2)$ is considered, the total differential of this function can be written as:

$$dy = \frac{\partial f}{\partial x_1}dx_1 + \frac{\partial f}{\partial x_2}dx_2 = MPP_1 dx_1 + MPP_2 dx_2 \tag{4.7}$$

The formal representation of the isoquant is achieved by considering the changes in x_1 and x_2 for which dy is equal to zero. (Hence, when dy is zero there are no changes in the production y which is in fact the characteristic feature of points on the isoquant for y). If dy in Eq. 4.7 is set equal to zero and solved with regard to dx_2/dx_1 the following result is generated:

$$\frac{dx_2}{dx_1} = -\frac{MPP_1}{MPP_2} = -MRS_{12} \tag{4.8}$$

which is in fact the ratio between changes in x_1 and x_2 when producing the constant product amount y. According to Eq. 4.8, the slope of the isoquant (dx_2/dx_1) can thus

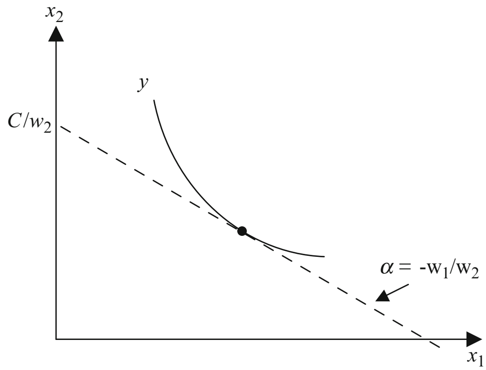


Fig. 4.2 Isoquant and isocost line

be expressed by the negative ratio between the marginal products for the two inputs MPP_1 and MPP_2. This ratio is called *M*arginal *R*ate of *T*echnical *S*ubstitution (*MRTS*), or simply *MRS* (*M*arginal *R*ate of *S*ubstitution). This is simply an expression of the amount of one input needed to compensate for a reduction in the other input under the condition that an unchanged amount of y is produced.

In Fig. 4.2, the isoquant for the production of y has been drawn. The minimisation of the costs C in Eq. 4.5 for the given production y can now be illustrated graphically, as shown in Fig. 4.2. According to Eq. 4.6, the result of a minimisation of C is that the isocost line is shifted as far to the south-west as possible, since the intersection point with the x_2 axis will then be placed as far down as possible. And as the input price w_2 is a constant, this will result in the lowest possible costs of C. At the same time, it is important to make sure that the amount y is produced, i.e. that production takes place somewhere on the isoquant for y. The optimal point is in fact the tangent point between the isocost line and the isoquant as shown in the figure. In this case, the slope of the isocost line ($-w_1/w_2$) is in fact equal to the slope of the isoquant ($-MPP_1/MPP_2 = -\mathrm{MRS}_{12}$) while at the same time producing y. This in fact corresponds to the condition (4.4) as derived previously.

In Fig. 4.1, the expansion path *ee* was drawn under the assumption that $-w_1/w_2 = -½$ which corresponds to the slope α of the three straight lines that are tangent to the three isoquants.

Example 4.1

Assume a production function $y = f(x_1, x_2) = 6x_1^{0.3}x_2^{0.5}$. The price of input x_1 is 8 ($w_1 = 8$), and the price of input x_2 is 12 ($w_2 = 12$). The Lagrange function L therefore equals:

$$L = 8x_1 + 12x_2 + \lambda\left(\mathrm{y} - 6x_1^{0.3}x_2^{0.5}\right).$$

Differentiating L with regard to x_1 and x_2 and setting the derivatives equal to zero will result in the following necessary conditions for an optimal production:

1. $8 = \lambda\, 1.8 x_1^{-0.7} x_2^{0.5}$
2. $12 = \lambda\, 3 x_1^{0.3} x_2^{-0.5}$

   Dividing 1. by 2. results in:

3. $\frac{8}{12} = \frac{1.8}{3} \frac{x_2}{x_1}$

   which can be written as:

$$\frac{x_2}{x_1} = \frac{24}{21.6}$$

Hence, to achieve optimal production in this example the inputs should be applied in the ratio 24/21.6. This means that a given product amount y is produced in the cheapest possible way by using x_2 and x_1 in the ratio 24/21.6. Hence, the points satisfying this condition are the points where the isocost line is tangent to an isoquant. As the collection of such points at the same time makes up the definition of the expansion path, then the expansion path in this example is given by the straight line through the zero point: $x_2 = \frac{24}{21.6} x_1$.

4.3 The Expansion Path and the Form of the Production Function

The expansion path in Fig. 4.1 is deliberately drawn as an arbitrary (non-linear) curve. The reason for this is that it is in fact not possible to say anything about the shape of the expansion path, unless the form of the production function is known.

If the expansion path constitutes a straight line through the origin, then the *production technology* is *homothetic* and in this case the production function $f(x_1, x_2)$ is called a *homothetic production function* (hence, the production function in Example 4.1. is a homothetic production function as it entails a linear expansion path). Hence, what makes a homothetic production technology special is that the variable inputs should always be used in the same ratio regardless of the level of production. Points on a straight expansion path through the zero point do in fact constitute a constant ratio between the inputs corresponding to the slope of the line.

If the production function is homothetic, the economic issues related to the adjustment of the production to changes in price ratios are simplified. If the product price p_y increases and production therefore should be expanded, there is no need to consider the ratio between the inputs when using a homothetic production technology. The inputs should simply be used in the same ratio as before. The same is true when reducing production in the case of falling prices.

A Cobb-Douglas production function has an expansion path which is a straight line through the origin. A Cobb-Douglas production function with two variable inputs has the form:

$$y = A x_1^{b_1} x_2^{b_2}.$$

Differentiating this production function with regard to x_1 and x_2, respectively, produces:

$$MPP_1 = Ab_1 x_1^{b_1 - 1} x_2^{b_2}$$

and

$$MPP_2 = Ab_2 x_1^{b_1} x_2^{b_2 - 1}$$

The equation for the expansion path is produced by using the general condition for the expansion path derived in Eq. 4.4. If MPP_1 is divided by MPP_2 and inserted in Eq. 4.4, the following result is generated:

$$\frac{w_1}{w_2} = \frac{b_1 x_2}{b_2 x_1},$$

which can also be expressed as:

$$x_2 = \frac{w_1}{w_2} \frac{b_2}{b_1} x_1,$$

which is the equation describing the expansion path as a straight line in the x_1–x_2-plane (see also Example 4.1).

Whether an assumption of production functions being homothetic is realistic or not will not be discussed here. It should merely be established that the precondition of a homothetic production technology demands that the optimal ratio between inputs does *not* depend on the scale of production. For instance, the optimal ratio between the consumption of the four inputs labour, acreage, fertiliser, and machinery used for the production of cereal crops is the same regardless of whether one is talking about a farm with 5 ha or 100 ha. This precondition can also be expressed as demanding that the optimal consumption of labour, machinery, and fertiliser per hectare will be the same regardless of the number of hectares cultivated. Concerning car manufacturing, the precondition of a homothetic production technology would imply that steel, labour, paint, fuel, plastic, tyres, etc. are used in the same proportion no matter how many cars are produced. Please take a moment to consider whether these assumptions are realistic. Could empirical observations provide a basis for the acceptance of such assumptions?

Homogeneous production functions are a particular class of homothetic functions. A homogeneous production function is characterised by the fact that – apart from being homothetic – it can be expressed as the production function:

$$f(tx_1, tx_2) = t^n f(x_1, x_2),$$

in which t is a positive number ($t > 0$) and n is *the degree of homogeneity*.

Illustrating that such a production function is in fact homothetic is easy. If the following is true for a given set of x's (x_1, x_2):

$$\frac{w_1}{w_2} = \frac{MPP_1}{MPP_2}$$

i.e. that production takes place on the expansion path, then – if the production function is homogeneous – the following will be true for another set of x's (tx_1, tx_2):

$$\frac{w_1}{w_2} = \frac{t^n MPP_1}{t^n MPP_2}$$

which can be reduced to:

$$\frac{w_1}{w_2} = \frac{MPP_1}{MPP_2}$$

If all inputs are increased by a factor t (movement along the line through the zero point), the isoquants along this line will then have the same slope $-MPP_1/MPP_2$, which corresponds to the expansion path being a straight line through the zero point. However, this is in fact the definition of a homothetic production function.

Apart from being homothetic, homogeneous production functions are special in the sense that if all inputs are multiplied by a factor t, then production increases by t^n independent of the present production level.

We will now have a look at the Cobb-Douglas production function as introduced above. As previously discussed, this function is homothetic. However, it is actually also homogeneous. Firstly, all inputs are multiplied by the factor t producing the following:

$$y = f(tx_1, tx_2) = A(tx_1)^{b_1}(tx_2)^{b_2}$$

which can be written as:

$$y = t^{(b_1+b_2)} A x_1^{b_1} x_2^{b_2} = t^n f(x_1, x_2).$$

Hence, a Cobb-Douglas production function is shown to be homogeneous of the degree ($b_1 + b_2$). Thus, if all inputs are doubled, then production will increase by a factor of $2^{(b1+b2)}$. If e.g. ($b_1 + b_2$) equals 1, the doubling of all inputs means that the production will actually be doubled.

A production function where the degree of homogeneity n is precisely 1 is homogeneous of degree one, or *linear homogeneous*. If a production function which includes all inputs (all inputs are variable) is linear homogeneous, then it is said to have *constant returns to scale*. This concept is derived from the observation that if the *scale* increases (all inputs increase with a given factor) for such production functions, then production increases with the same factor (discussed in further detail in Chap. 11).

Historically, the Cobb-Douglas production function has been much used as the functional form describing production – within both farming and industry. The popularity of this function is due to a number of mathematical advantages and advantages in connection with empirical analyses which will not be discussed further here. It must however be emphasised that this functional form demands a number of relatively stringent assumptions concerning the production technology. It is first and foremost the assumption that the expansion path is linear – i.e. that input – with given input prices – should always be used in the same ratio, regardless of the size of the production. In addition to this, there is the homogeneity assumption which demands that the degree of homogeneity is the same everywhere, i.e. globally.

As mentioned, the Cobb-Douglas production function is a homogeneous function with a degree of homogeneity n that equals the sum of the exponents b_1 and b_2. Other production functions are not necessarily homogeneous functions. For instance, a quadratic production function:

$$y = f(x_1, x_2) = a_0 + a_1 x_1 + a_2 x_2 - a_{11} x_1^2 - a_{22} x_2^2 + a_{12} x_1 x_2$$

is generally neither homogeneous nor homothetic (the reader is encouraged to find out under which preconditions the mentioned quadratic production function is in fact (1) homothetic or (2) homogeneous). However, for the given values of x_1 and x_2 it is of course possible to calculate how much the production y will rise if all inputs were increased by the same factor t.

Example 4.2

Presume that the parameters in the above mentioned quadratic production function have the values a_0 equal to 2, that a_1 and a_2 are both equal to 1, that a_{11} is equal to 0.10, that a_{22} is equal to 0.01, and that a_{12} is equal to 0.50. Based on this, calculate how much the production y increases when all inputs are increased by 10% compared to the present consumption of 1 unit of x_1 and 1 unit of x_2.

Initially, the production is:

$$y = 2 + 1 + 1 - 0.10 - 0.01 + 0.5 = 4.39$$

If all inputs are increased by 10% the production will be:

$$y = 2 + 1.1 + 1.1 - 0.121 - 0.0121 + 0.605 = 4.6719$$

which is an increase of (4.6719−4.3900)/4.3900 = 0.064, corresponding to 6.4%.

When, as in this case, production increases by a percentage that is smaller than the increase in all inputs (the factor t), this is referred to as *decreasing returns to scale*. If, on the other hand, production increases more than the increase in inputs, this is

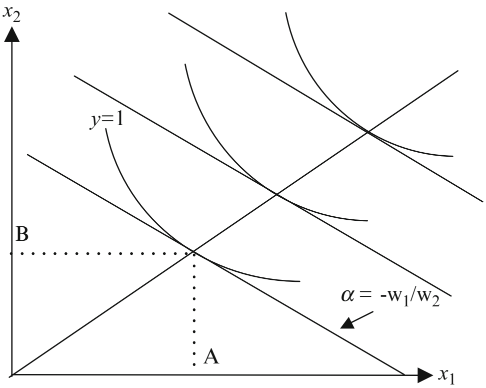


Fig. 4.3 Homothetic production function

referred to as *increasing returns to scale*. Finally, it is possible to talk about *constant returns to scale*, if production increases with the same percentage as (all) inputs.

If the effect of a 10% change in the amount of input is measured compared to another point of reference, this example will produce another production change percentage. (The reader is encouraged to calculate the effect of a 10% increase when the reference point is e.g. 5 units of x_1 and 5 units of x_2). A production function where this is the case is called a production function with a variable degree of homogeneity.

Homothetic production functions are, as mentioned, characterised by the optimal combination of inputs being constant, regardless of the level of the production. The reason is that the isoquants are parallel. Figure 4.3 shows a homothetic production function. Here it is illustrated that the optimal production of 1 unit of y takes place by the use of A units of x_1 and B units of x_2. As the production function is homothetic, the mentioned input prices demand that x_1 and x_2 must always be used in the ratio A/B.

Now presume that an input basket with A units of x_1 and B units of x_2 is created. With such a "basket" of inputs it is in fact possible to produce one unit of y. This basket is referred to as ⊗.

The production of y can now be illustrated in a figure similar to the one used for the analysis of one input. The input basket ⊗, which has just been created, can in fact be considered as being the input unit and, based on this, the production of y can be illustrated as shown in Fig. 4.4.

The concept of returns to scale in a multi-input context can now be illustrated graphically. As long as the application of input ⊗ is lower than P, returns to scale are increasing. The returns to scale are constant exactly at the input application P. And when the input application is greater than P, returns to scale are decreasing.

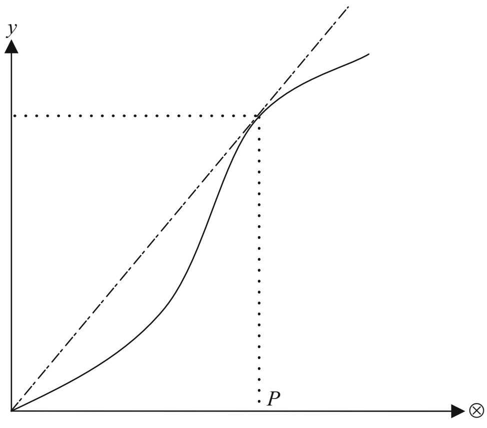


Fig. 4.4 Homothetic production function

4.4 Maximisation of Production Under Budget Constraint

The expansion path can also be derived as the solution to the following problem:

$$\max_{x_1, x_2}\{f(x_1, x_2)\} \tag{4.9a}$$

under the constraint that:

$$C = w_1 x_1 + w_2 x_2 \tag{4.9b}$$

The problem (4.9a)–(4.9b) consists of maximising the production Eq. 4.9a under the constraint that it is not possible to buy variable input for more than MU C (The budget constraint Eq. 4.9b. Here and in the following MU means Monetary Units). The solution is found by using the Lagrange method (see Chiang (1984) p. 372). The Lagrange function L is expressed as:

$$L = f(x_1, x_2) + \theta(\mathrm{C} \; - (w_1 x_1 + w_2 x_2)) \tag{4.10}$$

and maximised with regard to the two variables x_1 and x_2 as well as the Lagrange multiplier θ by taking the partial derivatives and setting them equal to zero. This produces the following three conditions for the optimal solution (4.11a)–(4.11b):

$$w_1 = MPP_1/\theta \tag{4.11a}$$

$$w_2 = MPP_2/\theta \tag{4.11b}$$

$$C = w_1x_1 + w_2x_2 \tag{4.11c}$$

Dividing Eq. 4.11a by Eq. 4.11b produces the necessary condition for the maximisation of Eq. 4.9a for the given C:

$$\frac{w_1}{w_2} = \frac{MPP_1}{MPP_2} \tag{4.12}$$

which turns out to be identical to the condition (4.4). Hence, the desire to minimise the costs for a given production or to maximise production for a given cost (budget) requires the use of the same criterion.

4.5 Profit Maximisation

The points on the expansion path are interesting as the company on the expansion path is in fact producing the given amount in the cheapest possible way (or producing the highest amount within the framework of a given budget constraint). The concept of "the expansion path" refers to the "path" along which to "expand" the production.[2]

If there are no constraints attached to the purchase or use of the two inputs x_1 and x_2, the rational producer will, in such a case, increase production by increasing the application of input along the expansion path. Presume that the optimal application of two inputs (corresponding to profit maximisation) corresponds to the point A in Fig. 4.5.

Point A is, as the other points on the expansion path, characterised by:

$$\frac{w_1}{w_2} = \frac{MPP_1}{MPP_2} \tag{4.13}$$

which can be also written as:

$$\frac{MPP_2}{w_2} = \frac{MPP_1}{w_1} \tag{4.14}$$

[2] Please note in this connection that the expansion path as presented here is a *stationary image* as, in reality, the (relative) input prices are presumed to be constant, and the production function is presumed to be unchanged. In the real world, an expansion of production will take time (it takes e.g. time to build a new building), and when the expansion at a later point in time has actually been carried out, then the prices w_1 and w_2 might have changed, and the production function $f(x_1, x_2)$ might also have changed due to the technological development.

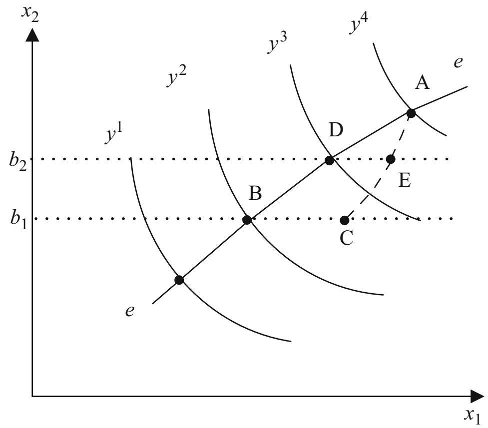


Fig. 4.5 Isoquants and profit maximisation

Furthermore, (as it will soon turn out) precisely at the profit maximising point A, the two fractions in Eq. 4.14 equal 1 divided by the price of output y, i.e. $1/p_y$. Hence, what makes the profit maximising point A special is that:

$$\frac{MPP_2}{w_2} = \frac{MPP_1}{w_1} = \frac{1}{p_y} \tag{4.15}$$

or that:

$$\frac{VMP_2}{w_2} = \frac{VMP_1}{w_1} = 1 \tag{4.16}$$

in which VMP_i (the value of the marginal product for input i) is $MPP_i p_y$.

The criterion for profit maximisation with two inputs in Eq. 4.16 can be derived by maximising the profit as the function of the two inputs:

$$\max_{x_1, x_2} \{f(x_1, x_2)p_y - w_1 x_1 - w_2 x_2\} \tag{4.17}$$

Differentiating the profit in Eq. 4.17 with regard to x_1 and x_2 and setting the partial derivatives equal to zero results in the following *conditions for profit maximisation*:

$$w_1 = MPP_1 p_y (\equiv VMP_1) \tag{4.18a}$$

$$w_2 = MPP_2 p_y (\equiv VMP_2) \tag{4.18b}$$

which in fact corresponds to the criteria in Eqs. 4.15 and 4.16. Hence, the previously derived result regarding one input is (naturally) also true for two inputs, i.e. so that the value of the last unit (VMP_i) corresponds to the price w_i of this unit

($i = 1, 2$). This result can be easily generalised to cover more inputs so that the *criterion for profit maximisation with n variable input* is:

$$\frac{VMP_1}{w_1} = \frac{VMP_2}{w_2} = \ldots = \frac{VMP_n}{w_n} = 1 \tag{4.19}$$

Generally, it is presumed that producers maximise profit and in so doing in fact seek to satisfy the condition (4.19) when purchasing or adding variable input. However, there might be situations in which producers cannot, or do not wish to maximise profit. This might e.g. be the case under budget constraints where the producer does not have sufficient funds to buy the amount of variable input needed to satisfy the condition (4.19).

In the previous it has been shown that under budget constraints inputs should be combined to satisfy Eq. 4.4. Please note that Eq. 4.4 can also be written as:

$$\frac{MPP_1 p_y}{w_1} = \frac{MPP_2 p_y}{w_2} \tag{4.4a}$$

as multiplication with a constant p_y on both sides of the equal sign does not change the ratio. Therefore, Eq. 4.4a can also be written as:

$$\frac{VMP_1}{w_1} = \frac{VMP_2}{w_2} \tag{4.4b}$$

Comparing Eq. 4.4b with Eq. 4.16, which expresses the criterion for profit maximisation, shows that the characteristic feature of the profit maximum, compared to other points along the expansion line, is that precisely at the point of profit maximum, the ratio of VMP_i/w_i equals 1. However, what is the ratio for the points on the expansion path that are placed *before* the profit maximum?

The answer is that under the assumption of diminishing marginal productivity, the ratio stated is greater than 1. The reason is that when MPP_i is diminishing with increasing x, then the numerator of the fraction in Eq. 4.4a is higher than for the corresponding fractions in Eq. 4.19 when the input supply is smaller than that which corresponds to the profit maximum. A more general criterion for the combination of input is therefore:

$$\frac{VMP_1}{w_1} = \frac{VMP_2}{w_2} = \cdots = \frac{VMP_n}{w_n} \geq 1 \tag{4.20}$$

as the profit maximum thereby represents the special case in which the ratio stated is equal to 1.

The optimisation criterion in Eq. 4.20 is one of the key results in the theory of production economics and should therefore be pointed out here. Expressed in words, the criterion could be described as follows:

Key Result

> Multiple variable inputs must always be combined so that the ratio between the value of the marginal product and the input price is the same for all inputs. If there are no budget constraints or other restrictions, the supply of all inputs should be increased to the point where the ratio between the value of the marginal product and the input price equals 1.

4.6 The Pseudo Scale Line and Optimisation with Fixed Inputs

What happens if one of the inputs that used to be variable becomes fixed? Presume e.g. that input x_2, which used to be a variable input, becomes a fixed input because it is – for some reason or another – only available in a given fixed amount b_1, which is less than the optimal amount when x_2 was a variable input. How can the input supply, and thereby the production, be adjusted optimally? Should the adjustment take place along the "old" expansion path, i.e. should production be reduced to point B in Fig. 4.5? (Point A in Fig. 4.5 illustrates the profit-maximising production when both x_1 and x_2 are variable inputs).

No, point B is in fact not optimal. This is easy to see because under the assumption of a diminishing marginal product everywhere, the following is true for point B:

$$\frac{MPP_1 p_y}{w_1} = \frac{VMP_1}{w_1} > 1 \tag{4.21}$$

as MPP_1 is larger at point B than at point A (also cf. (4.20))

The inequality in Eq. 4.21 entails that it pays to increase the supply of x_1 at point B. And there is nothing to prevent that from being done, as x_1 is a variable input. By how much should the application of x_1 be increased? Well, according to the general rules for the optimisation of one variable input, supply should be increased as long as the value of the marginal product (VMP_1) is greater than the input price (w_1) and stopped when the two expressions are the same, i.e. when:

$$\frac{VMP_1}{w_1} = 1 \tag{4.22}$$

which is true for a point to the right of B, e.g. at point C in Fig. 4.5.

Now presume instead that the producer had a fixed amount b_2 of the input x_2 at his/her disposal. Applying the same arguments as just used shows that point D on the original expansion path is not optimal. It would be optimal to increase the supply of x_1, for instance to point E where the value of the marginal product of input x_1 is equal to the input price w_1 corresponding to Eq. 4.22.

If the same analysis is carried out for all possible fixed levels of the input x_2, Fig. 4.5 illustrates a curve of optimal points through the points C, E, and A (as A also satisfies the condition (4.22) cf. (4.20)). This curve is called *the pseudo scale line*. *Hence, the pseudo scale line describes the relationship between the various levels of a fixed input x_2 and the corresponding optimal application of a variable input x_1.*

This pseudo scale line will prove useful in Chap. 10 where we will be looking at the optimisation of production under restrictions.

4.7 Substitution Between Inputs

The fact that two inputs can replace each other in connection with the production of a given amount of output has been graphically illustrated by means of the so-called isoquants in the above. The shape of such isoquants is an indication of how easy it is to replace one input for another. In Fig. 4.6 below, three different degrees of input substitution are illustrated.

Part A in Fig. 4.6 shows a production with full substitution between the two inputs and a *constant substitution rate* (dx_2/dx_1 ($-MRS_{12}$) – see Eq. 4.8) corresponding to the slope of the straight line. Part B and C illustrate a *decreasing substitution rate* as an increasing amount of one input replaces a continuously diminishing amount of the other input. The substitution possibility is, however, larger in B than in C. Finally, in part D there is no substitution possibility (more of one input *cannot* replace part of the other input, if a product amount corresponding to the isoquant should still be produced).

The mathematical expression for the substitution rate (MRS) can be used to describe the degree of substitution. However, the so-called *elasticity of substitution* is often used, as the elasticity of substitution is a unit-free concept which – as is always the case with elasticities – expresses the relative change in one expression divided by the relative change in another expression. *The input elasticity of substitution* (ε_{sh}) originally proposed by Earl O. Heady (1952) can be approximated for small changes (Δx) by:

$$\varepsilon_{Sh} = \frac{\Delta x_2}{x_2} \Big/ \frac{\Delta x_1}{x_1} \tag{4.23}$$

but is more formally *defined* as:

$$\varepsilon_{Sh} = \frac{dx_2}{x_2} \Big/ \frac{dx_1}{x_1} = \frac{dx_2}{dx_1} \frac{x_1}{x_2} \tag{4.24}$$

i.e. as the slope of the isoquant multiplied by the ratio between x_1 and x_2 at the point where the elasticity is measured.

As can be seen, the elasticity of substitution has the same sign as the slope of the isoquant (i.e. negative at the relevant part of the isoquant). When talking about the value of the elasticity of substitution it is, however, common to refer to its absolute value. This is also the case in the following.

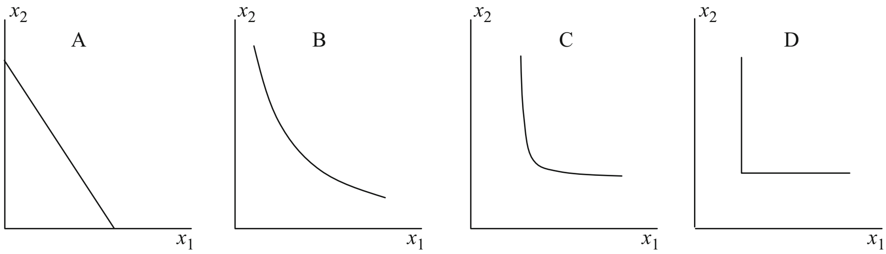


Fig. 4.6 Alternative shapes of isoquants

The substitution elasticity will normally depend on the position on the isoquant. The substitution elasticity of the linear isoquant (A) in Fig. 4.6 increases e.g. from 0 to infinity when the amount of x_1 is increased from 0 to the maximum amount (where the isoquant intersects the x_1 axis).

Certain production functions have isoquants that are characterised by the substitution elasticity being constant. This is e.g. true for the Cobb-Douglas production function where the slope (dx_2/dx_1) of the isoquant (as shown in Sect. 4.3) is equal to $-b_1x_2/b_2x_1$. Multiplying this by x_1/x_2 (see 4.24) results in the substitution elasticity $\varepsilon_s = -b_1/b_2$ which is constant, i.e. independent of the x's. The isoquant in part B in Fig. 4.6 could be illustrating such an isoquant.

Generally speaking, the substitution elasticity is a (local) expression of how well the observed inputs replace each other. A high substitution elasticity ($|\varepsilon_s| > 1$) is an indication that it will be possible to save a relatively large amount of one input by adding a relatively small extra amount of the other input. A small substitution elasticity ($|\varepsilon_s| < 1$) is an indication that it will only be possible to save a relatively small amount of one input even though a relatively large extra amount of the other input is added.

The substitution elasticity is a well-defined concept when talking about production functions with only two inputs. However, if there are three or more inputs the definition is not entirely unambiguous, as the substitution elasticity (between two inputs) will often depend on how much has been added of the other input beforehand. This issue will not be discussed any further here. Please refer to more advanced textbooks (see e.g. Chambers 1988, p. 27ff.).

References

Chambers, R. G. (1988). *Applied production analysis: A dual approach*. New York: Cambridge University Press.

Chiang, A. C. (1984). *Fundamental methods of mathematical economics* (3rd ed.). Singapore: McGraw-Hill Book Company.

Heady, E. O. (1952). *Economics of Agricultural Production and Resource Use*. New York: Prentice-Hall.

Costs

5

5.1 One Variable Input

Costs are the monetary value of input used over a period of time. A company's costs can be derived from the production function.

The point of reference is a production function with one variable input x_1 and an output y, as shown in the left hand side of Fig. 5.1. If you imagine this figure removed from the paper, lifted up, and then put down again with the front side down and turned 90° clockwise, then you'll get the figure – the cost function – in the right hand side of Fig. 5.1.

The curve in the right hand side of Fig. 5.1 is not, of course, an entirely correct cost function. Costs are measured in monetary terms (MU), and the unit of measurement on the vertical axis on the figure in the right hand side is not MU, but units of input x_1. However, if the units on the vertical axis are multiplied by the price w_1 of x_1, and if x_1 is furthermore measured in units having the exact price of $w_1 = 1$, then the figure to the right does in fact measure the variable costs ($VC(y) = w_1 x_1(y)$) as a function of the production y, as the applied input amount $x_1$is rendered as a function of the production y.

The (*dual*) relationship between the production function and the cost function illustrated here is essential to understanding the modern approach to the estimation of the production function and other production-related relationships. It is important to understand that with knowledge of the production function it is possible to determine the cost function (move from left to right in Fig. 5.1), whilst conversely, it is also possible – with knowledge of the cost function – to determine the production function (move from right to left in Fig. 5.1). This so-called theory of duality shall not be discussed in further detail here, as this is the subject of descriptive production economics (see e.g. Chambers (1988) for a discussion of duality in production theory).

The production function

$$y = f(x_1)$$

S. Rasmussen, *Production Economics*, Springer Texts in Business and Economics,
DOI 10.1007/978-3-642-30200-8_5,

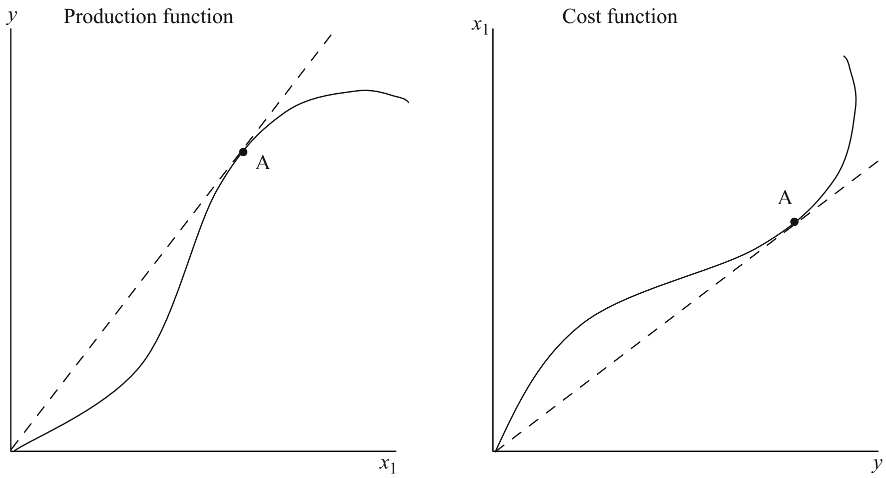


Fig. 5.1 Cost function as a mirror reflection (dual) of the production function

in Fig. 5.1 expresses the production as a function of the variable input x_1. Normally, multiple inputs are used which can be explicitly expressed as $y = f(x_1|x_2, \ldots, x_\mathrm{n})$, where the inputs $x_2, \ldots, x_\mathrm{n}$ are fixed inputs. Similarly, the following:

$$VC(y) = w_1 x_1(y) \tag{5.1}$$

solely measures the *variable costs*. The use of fixed inputs also includes costs, i.e. *fixed costs* expressed as:

$$FC = w_2 x_2 + \ldots + w_n x_n \tag{5.2}$$

As the amounts $x_2, \ldots x_n$ are presumed to be fixed, and as the input prices $w_2, \ldots,$ w_n are presumed to be given (and, hence, fixed), FC is a constant – independent of the production y.

Adding up the variable and fixed costs gives the *total costs TC* expressed as:

$$TC(y) = VC(y) + FC \tag{5.3}$$

The mentioned cost concepts are illustrated graphically in Fig. 5.2.

The average costs can be directly defined as:

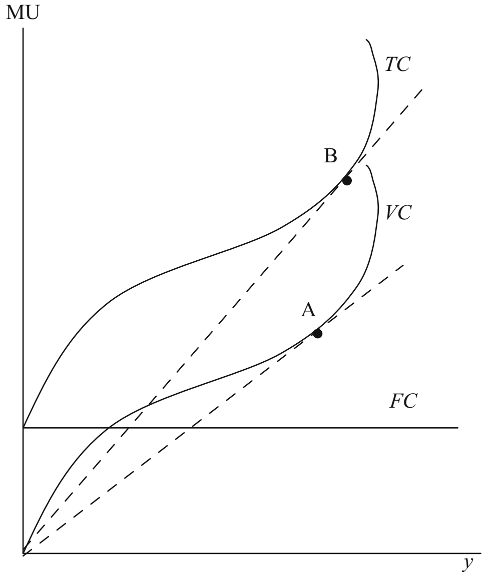
Average variable costs:

$$AVC(y) = VC(y)/y \tag{5.4}$$

Average fixed costs:

$$AFC(y) = FC/y \tag{5.5}$$

Fig. 5.2 Fixed, variable, and total costs



Average total costs:

$$ATC(y) = TC(y)/y \tag{5.6}$$

Graphically, the average variable costs equal the slope of a straight line through the zero point up to the *VC* curve (see Fig. 5.2). Hence, the lowest variable average costs are found at point *A* where the slope of the straight line through the zero point to the variable cost curve is lowest. The average fixed costs equal the slope of a straight line through the zero point up to the *FC* curve. And, finally, the average total costs equal the slope of a straight line through the zero point up to the *TC* curve. Hence, the lowest average total costs are found at point *B*.

The *marginal costs* (the costs of producing one additional unit of *y*) are defined as:

$$MC(y) = \frac{\partial TC}{\partial y} = \frac{\partial VC}{\partial y} \tag{5.7}$$

which corresponds graphically to the slope of the cost curve – either the total costs curve or the variable costs curve (for a given value of *y*, the slope of the two curves is the same (see Fig. 5.2)). Figure 5.2 furthermore shows that the marginal costs equal the average variable costs precisely where these are at their lowest (point *A*), and that the marginal costs equal the average total costs precisely where these are at their lowest (point *B*).

The relationships mentioned here are graphically illustrated in Fig. 5.3 where the curves for the average and marginal costs are shown in the lower part of the figure.

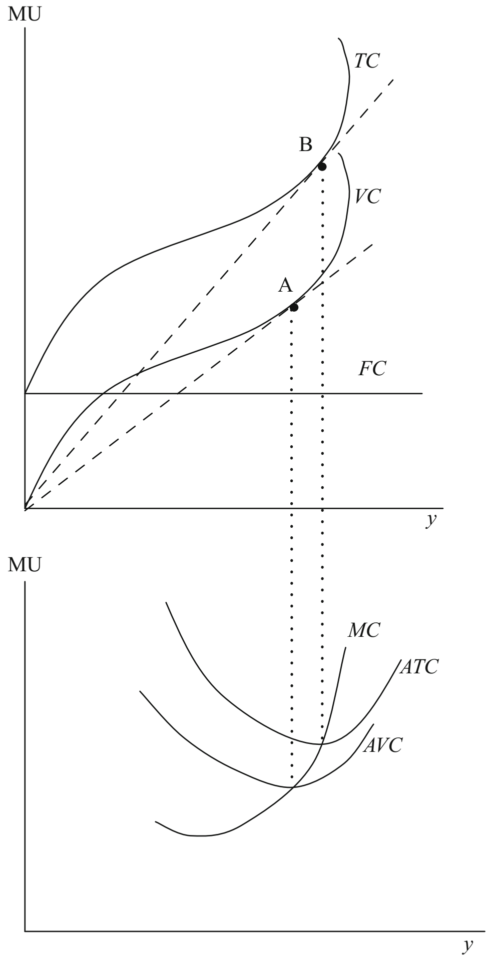


Fig. 5.3 Average costs and marginal costs

5.2 Multiple Variable Inputs

With multiple variable inputs, the cost function cannot be directly derived from the production function as illustrated in Fig. 5.1. The costs (or rather, the variable costs) will not only depend on the produced amount y but also on the combination of variable inputs used in the production. This again will depend on the prices of the variable inputs.

In Chap. 4, it was demonstrated that the optimal combination of inputs is found on the expansion path. With two variable inputs, the problem was formulated as (see Eqs. 4.1a and 4.1b):

$$\min_{x_1, x_2}\{w_1x_1 + w_2x_2\}$$

under the constraint that:

$$y = f(x_1, x_2)$$

and the criteria for the optimal combination of the two inputs (the expansion path) were derived based on this.

The same method can be used to determine the cost function, as the cost of the production of y with two (or more) inputs is defined as being the lowest possible cost by which the amount y can be produced. With two variable inputs, this results in the following formal *definition of the cost function*:

$$VC(y, w_1, w_2) = \min_{x_1, x_2}\{w_1x_1 + w_2x_2 | y = f(x_1, x_2 | x_3, \ldots, x_n)\} \tag{5.8}$$

The variable costs are now a function of both the produced amount y and of the prices of the variable inputs. The reason that it is the input prices and not the amount of input that are the arguments in the cost function is that it is the input prices that determine the optimal combination of input (the expansion path).

The definition of the variable costs in Eq. 5.8 can be directly generalised to cover more (k) variable inputs, so that the general definition of the variable cost function is:

$$VC(y, w_1, \ldots, w_k) = \min_{x_1, \ldots, x_k}\{w_1x_1 + \ldots + w_kx_k \, | y = f(x_1, \ldots, x_k \, | x_{k+1}, \ldots, x_n)\} \tag{5.9}$$

Example 5.1

In Chap. 4 (Sect. 4.3), the equation for the expansion path for a Cobb-Douglas production function was shown to be:

$$x_2 = \frac{w_1}{w_2} \frac{b_2}{b_1} x_1. \tag{5.10}$$

Introducing the expression for x_2 in the Cobb-Douglas production function and solving it for x_1 yields:

$$x_1 = \left(\frac{y}{A}\right)^{1/(b_1+b_2)} \left(\frac{b_1w_2}{b_2w_1}\right)^{b_2/(b_1+b_2)} \tag{5.11}$$

and similarly for x_2:

$$x_2 = \left(\frac{y}{A}\right)^{1/(b_1+b_2)} \left(\frac{b_2 w_1}{b_1 w_2}\right)^{b_1/(b_1+b_2)} \tag{5.12}$$

If these expressions for x_1 and x_2 are inserted in the formula for the calculation of the variable costs $VC = w_1x_1 + w_2x_2$ the following cost function is generated:

$$VC(y,w_1,w_2) = \left(yA^{-1}w_1^{b_1}w_2^{b_2}\right)^{1/(b_1+b_2)} \left(\left(\frac{b_1}{b_2}\right)^{b_2/(b_1+b_2)} + \left(\frac{b_2}{b_1}\right)^{b_1/(b_1+b_2)}\right) \tag{5.13}$$

which in fact expresses the variable costs as the function of the production y and the input prices w_1 and w_2.

Previously, in Chap. 4 (Sect. 4.3), it was mentioned that when a production function is homothetic, then the expansion path is a straight line through the zero point, and the optimal ratio between the two inputs x_1 and x_2 is thus constant. This means that when the input prices are given, the ratio between the two inputs (and thereby the expansion path) will also be given, and the costs will subsequently just be a function of how far along the expansion path one moves. This means that the cost function in this case is *separable* as it can be expressed as:

$$VC(y, w_1, w_2) = VC[g(y), h(w_1, w_2)] \tag{5.14}$$

A Cobb-Douglas production function is homothetic, and therefore Eq. 5.13 has the form Eq. 5.14, where:

$$g(y) = y^{1/(b_1+b_2)} \tag{5.15}$$

and $h(w_1, w_2)$ is the remainder of Eq. 5.13. The function $h(w_1, w_2)$ determines on which expansion path the production takes place, while $g(y)$ determines how far along the expansion path to move to produce y.

For given values of the input prices $w_1, \ldots, w_k$, the variable costs are solely a function of the production y, and the cost concepts that have been developed for a variable input (Eqs. 5.3–5.7) can therefore be directly applied to productions using multiple inputs. This is also true for the graphical illustrations in Fig. 5.3, which in turn is true for productions that are based on multiple inputs.

Example 5.2

In Example 4.1, it was demonstrated that the lowest costs of production of y are achieved by combining x_2 and x_1 in the ratio 24:21.6. The table below shows seven combinations of x_1 and x_2 and the corresponding costs C with input prices as outlined in Example 4.1. The production y is furthermore calculated by introducing the outlined values of x_1 and x_2 in the production function from Example 4.1.

		Costs	Production	Approximated marginal costs
x_1	x_2	C	y	$\Delta C/\Delta y$
2	2.20	42.42	10.96	
				5.22
4	4.40	84.84	19.08	
				5.80
6	6.61	127.27	26.40	
				6.21
8	8.81	169.69	33.23	
				6.53
10	11.01	212.11	39.72	
				6.80
12	13.21	254.53	45.96	
				7.03
14	15.41	296.95	51.99	

There is no actual cost function, and the marginal costs cannot therefore be directly calculated. The marginal costs are therefore approximated by the calculation of *incremental costs*, $\Delta C/\Delta y$. The marginal costs that are estimated in this way are an approximated expression for the marginal costs in the centre of the interval, i.e. that the marginal cost of 5.22 is an expression of the marginal cost when y is 15.05, and 5.80 is an expression of the marginal cost when y is 22.74, etc.

Based on the cost function, it is possible to develop an *alternative criterion for profit maximisation* (compare with Eqs. 4.17–4.19).

The alternative criterion for profit maximisation (with one output and any number of inputs) is derived by maximising the following expression for profit:

$$\max_{y} \{yp_y - VC(y, w_1, \ldots, w_k) - FC)\} \tag{5.16}$$

The formulation in Eq. 5.16 presupposes that the producer is the price taker, i.e. that the product price p_y is independent of the produced amount of y. (In Sect. 13.4,

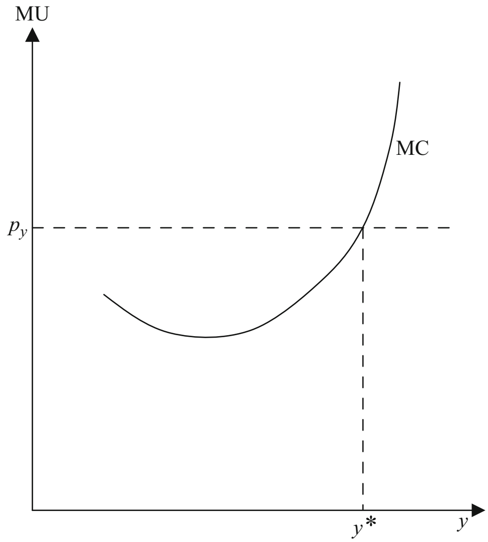


Fig. 5.4 Determination of optimal production

a similar condition is derived with the product price being dependent on the production y).

Differentiating the profit in Eq. 5.16 with regard to y and setting the derivative equal to zero yields:

$$p_y - MC(y) = 0 \tag{5.17}$$

or:

$$p_y = MC(y) \tag{5.18}$$

in which $MC(y)$ are the marginal costs defined in Eq. 5.7. The criterion (5.18) states that optimal production takes place when the output price is *equal to the marginal costs*.

The criterion is illustrated graphically in Fig. 5.4, in which y^* represents the optimal production. The shape of the marginal cost curve as a progressively rising curve has previously been derived in Fig. 5.3.

The criterion for profit maximisation in Eq. 5.18 generates the same result as with profit maximisation from the input side (see Eq. 4.19). Please note in this connection that the use of the cost function in Eq. 5.16 entails that a decision should have already been made as to how (with what combination of inputs) a given amount of output should be produced. The maximisation therefore only refers to the production y.

The optimisation criterion in Eq. 5.18 is one of the key results in the theory of production economics and should therefore be pointed out here. Expressed in words, the criterion could be described as follows:

Key Result

▶ Producers who want to maximise profit should continue to expand production as long as the marginal costs are lower than the product price, and halt further expansion of production at the exact point where the marginal costs are equal to the product price.

Please note that the maximised function:

$$\pi\left(p_y, w_1, \ldots, w_k\right) = \max_{y} \{yp_y - VC(y, w_1, \ldots, w_k) - FC\} \tag{5.19}$$

is referred to as the *profit function*.

Example 5.3

Example 5.2 included an example of the calculation of the marginal costs. If the product price e.g. is MU 5.80, then it is optimal to produce somewhere between 19 and 26 product units, as the marginal cost in this interval is precisely MU 5.80. If the product price increases to MU 7 per unit, it will be profitable to expand production to between 46 and 52 units of y, when the marginal cost is around MU 7.

5.3 Short and Long Run Costs

The cost curves previously shown in Fig. 5.3 are expressions of the costs in the *short run*. *Short run* means that part of the input factors are fixed factors giving rise to fixed costs (FC).

A company's production plant can often be considered as a fixed input factor in the short run. This would be e.g. buildings, machinery, and land. In the short run, these fixed assets entail fixed costs, as illustrated by FC in Fig. 5.3. The production with precisely such fixed assets is reflected by the variable cost curve VC in Fig. 5.3.

In the *long run*, the nature of the fixed input factors changes. In the long run, it is possible to change the company's fixed assets, thereby making the factors, which were previously fixed factors, variable input factors. The building facility, which was previously a fixed factor, can in the long run be adjusted with regard to size, as it is possible to invest in a new and possibly better building, or to expand the existing building. It is also possible to refrain from erecting a new building, when the existing one is run down.

These conditions are outlined in Fig. 5.5, in which the curve TC_1 and FC_1 correspond to the original cost curves in the upper part of Fig. 5.3.

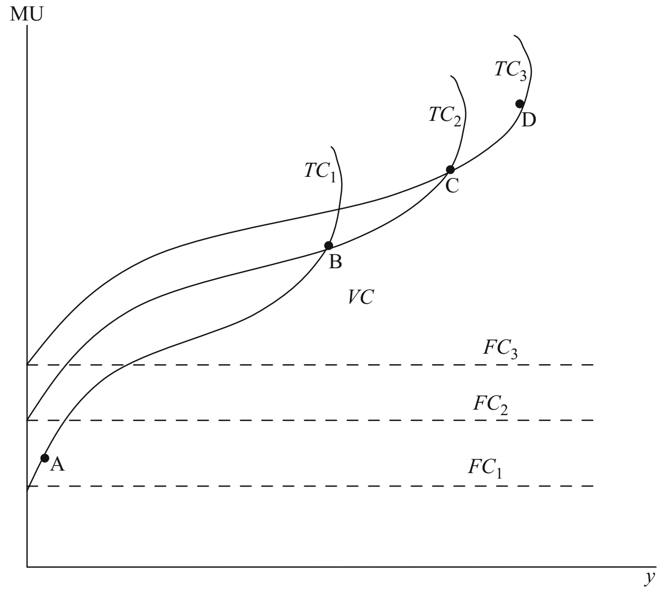


Fig. 5.5 Short run costs at different plant sizes

In the long run, it is possible to erect larger buildings. Building 2 has higher fixed costs (FC_2) but can produce larger amounts of output for the given amount of variable input, as illustrated by the cost curve TC_2.

Another possibility is to erect an even larger building – building 3 – which has even higher fixed costs (FC_3) but which can produce even greater quantities of the product for given amounts of variable input.

If you imagine that the size of buildings can be varied continuously, then the long run costs can be illustrated in a figure, in which the cost curve contains the possibility of varying the size of the building. Such a curve is illustrated in the upper part of Fig. 5.6, where the points A, B, C, and D correspond to the points with the same designation in Fig. 5.5.

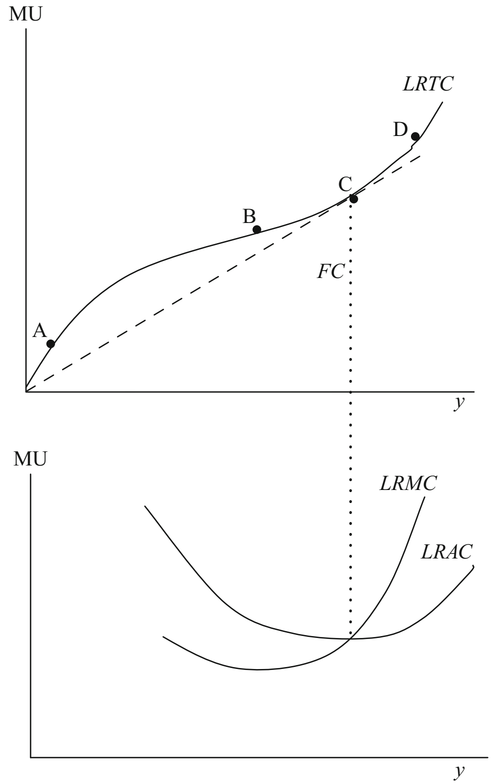
In the lower part of Fig. 5.6, the corresponding curves have been plotted for the long run average costs (*LRAC*) and the long run marginal costs (*LRMC*).

5.4 Calculation of Costs in Practice

As stated in the beginning of this chapter, costs are defined as the monetary value of input use over a period of time. As any monetary value is the product of quantity and price, the calculation of costs in practice involves two problems: the estimation of input quantities and the estimation of input prices.

For variable inputs traded at market prices, the calculation of costs is straight forward. If one decides to buy and use q units of an input which has a market price of w, then the cost is qw. But what if the firm already has the input in stock, because

Fig. 5.6 Long run costs



it has been bought at an earlier date? In this case, alternative prices may be used to estimate costs: (1) The original purchase price (the price at which the input was originally bought), (2) the present (actual) purchase price, (3) the present (actual) selling price, (4) other "prices" (for instance the internal value of the input).

From an *accounting perspective*, the obvious choice is to use the original purchase price. But this again depends on the *accounting principle* used. If the accounting principles are based on actual payments, then the original purchase price is the relevant price to use. However, if the accounting principles are based on the *replacement principle*, then the cost of using input in stock is the expenditure of replacing the input taken out of the stock, and the present (actual) purchase price would then be the relevant price to use.[1]

From an *economic* (versus accounting) *perspective*, costs should be estimated according to the *opportunity cost principle,* which means that costs are the value of missed opportunities. If the missed opportunity is to sell the input, then the actual selling price would be the relevant price to use when estimating costs. If the missed opportunity

[1] The accounting principles are further discussed in the Appendix on Profit concepts.

is to carry out production A, instead of production B, then the relevant cost of using the input in production A is the profit forgone by not using the input in production B.

Costs estimated using the accounting principle are also called explicit costs. *Explicit costs* are those costs that involve actual payment to other parties. Costs estimated according to the opportunity cost principle are also called *implicit costs*. *Implicit costs* represent the value of forgone opportunities, but do not involve actual cash payment.

In general, costs can be calculated according to the two principles: (1) *The opportunity cost principle* (implicit costs) and (2) *The accounting principle* (explicit costs) as follows:

1. The opportunity cost principle
   Calculation of the costs according to the opportunity cost principle is based on the alternative usage of the production factors. According to the opportunity cost principle, the costs are equal to *the earnings lost (lost opportunity) by not using the production factors in question in the best alternative way*. The opportunity cost principle is the key basis for all economic planning (a cost concept pointing to the future).
2. The accounting principle
   Calculation of the costs according to the accounting principle is based on *re-acquisition* of the production factors. Costs are calculated according to the accounting principle as the amount that should be used to reacquire the production factors used – or rather, the amount that should be used to restore the original the situation. The accounting principle is used in connection with the calculation of a financial profit and is, as such, directed towards the past ("history writing"). The accounting principle and its alternative versions are further discussed in the appendix *Profit concepts*.

The following overview provides some examples to illustrate these two principles:

	Costs	
Production factor	Opportunity cost principle	Accounting principle
Machine	Lost revenue by not letting the machine	Repair/maintenance and depreciation
Labour	Lost revenue by not using the labour in an alternative way, e.g. wage in connection with paid work	Food and beverages. But what about "depreciation"??
Buildings	Could they be used for something else?? If not, then the (implicit) cost is zero!	Repair/maintenance and depreciation
Land	Revenue in connection with leasing out the land. Gross margin in connection with usage for other crops	Are there any costs??
Purchased raw material	Purchase price	Purchase price
Fertiliser in stock	Lost revenue by not using the fertiliser for another crop	Purchase price in connection with refilling the stock
Livestock	Lost revenue by not selling the animal and depositing the money in the bank at r $\times$ 100% in interest	Fodder, veterinary service, "depreciation" (change of value)

These descriptions should be looked upon as examples only. Regarding the opportunity costs, the *best alternative* could, after all, vary from one situation to the other.

It should be noted that the fixed input factors, per definition, are input factors of an amount which cannot (or will not as it is undesirable) be varied over the planning period under consideration. Such (fixed) input factors which you cannot (or do not want to) sell will, therefore, per definition have zero opportunity costs. This is why the costs of such fixed factors are normally disregarded in connection with planning – exactly because the alternative cost is zero!

It should, however, be noted that even in the case where the opportunity cost for the company as a whole is zero, then there could, from an opportunity perspective, be costs in connection with the usage of the production factor in question in a given production. If the company, for instance, has several (alternative) production branches, then the usage of a production factor in one of the production branches will result in an (opportunity) cost if the same production factor could have been used in another production branch. If e.g. land is a fixed factor for the company as a whole (and the opportunity cost therefore is zero), then there are still costs in connection with the usage of the land for growing barley, as it could alternatively have been used for growing wheat. When calculating the costs of growing barley, the costs of land should therefore be included as the amount (gross margin) which could have been earned by growing wheat instead.

Reference

Chambers, R. G. (1988). *Applied production analysis: A dual approach*. New York: Cambridge University Press.

Productivity, Efficiency and Technological Changes

6

6.1 Introduction

The description of the production within an industry is often based on empirical data. In Denmark, there is an abundance of data for the description of production within farming. On the micro-economic level, this would be, for example, notes and financial accounts from the individual farms, and on an industry level it would be various kinds of statistical information describing production, factor consumption, prices etc.

The development in production and input factor consumption over time is often of considerable interest. The description of the increase or decrease in production can be presented in various ways and can e.g. be related to the factor consumption. An increase (or decrease) in production can be interesting in itself. However, changes in the production will often be compared to changes in the factor consumption. If production increases more than the factor consumption then this is referred to as increased productivity. Other concepts are also used to discuss and evaluate changes in production and factor consumption. Concepts such as productivity, efficiency, and technological changes are often used. However, these concepts are often used without the speaker being entirely aware of their precise meaning.

This chapter examines how these concepts are defined and how they are related. It will also examine why it may be interesting to describe these measures and their development over time.

6.2 Definitions

6.2.1 Productivity

Productivity can be briefly defined as production (output) divided by input. In a production where only one input x is used to produce one output y, the description is simple, as productivity will then be y/x, i.e.:

S. Rasmussen, *Production Economics*, Springer Texts in Business and Economics,
DOI 10.1007/978-3-642-30200-8_6,

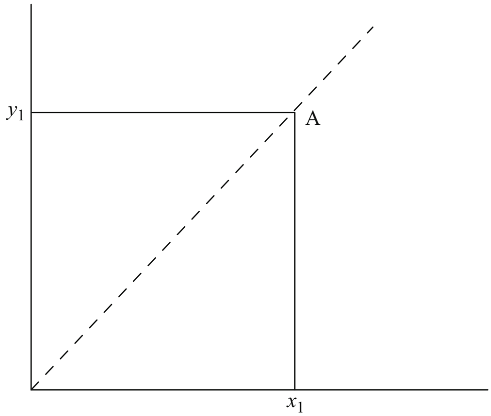


Fig. 6.1 Illustration of productivity

$$\text{Productivity} = \mathrm{P} = y/x \tag{6.1}$$

If production and factor consumption in period t is y_t and x_t, respectively, and in period $t + 1$ is y_{t+1} and x_{t+1}, respectively, then *the change in productivity* from period t to period $t + 1$ equals:

$$\text{Change in productivity} = dP = \left(\frac{\frac{y_{t+1}}{x_{t+1}} - \frac{y_t}{x_t}}{\frac{y_t}{x_t}}\right) = \left(\frac{\frac{y_{t+1}}{x_{t+1}}}{\frac{y_t}{x_t}} - 1\right) = \left(\frac{y_{t+1}}{y_t}\frac{x_t}{x_{t+1}} - 1\right) \tag{6.2}$$

The last parenthesis in Eq. 6.2 illustrates that productivity increases over time can be achieved either by an increase in the production y, or by a decrease in the consumption of input x.

Example 6.1

The consumption of input is 30 units in the year 2006 and 35 units in the year 2007. The production is 140 units in the year 2006 and 180 units in the year 2007. The increase in productivity from the year 2006–2007 therefore equals (180/140)(30/35)−1 = 0.102, or 10.2 %.

Productivity can be illustrated graphically as the slope of the line through point A in Fig. 6.1, in which x_1 is the amount of input and y_1 is the amount of output.

If the production (of one output y) takes place by the use of multiple inputs ($x_1 \ldots x_n$), multiple measurements of productivity can in principle be calculated. Hence, for each of the n inputs it is possible to calculate a *partial measurement of productivity* by simply introducing one of those n inputs in the above formulas. The cereal crop yield per hectare and the number of pigs per sow are examples of such partial measurement of productivity within farming.

It is also possible to aggregate all inputs using a formula to calculate an input index. An *input index* is a number expressing the total consumption of input. A well-known index is the so-called Laspeyre's quantity index which is calculated as:

$$QI = QI_L^t = \frac{\sum_{k=1}^{n} w_{tk} x_{t+1,k}}{\sum_{k=1}^{n} w_{tk} x_{tk}}$$

in which x_{tk} is the consumption of input k in the period t, w_{th} is the input price of input k in period t, and QI_L^t is the Laspeyre's quantity index of consumption of all inputs in the period t + 1 when the consumption in the period t is set equal to 1. There are many other methods for calculating quantity indices, but it will be too comprehensive to discuss them here (If you want to know more about indices, please refer to the vast literature on index theory, for instance Balk 1998).

If the input index is called QI, then the so-called Total Factor Productivity (TFP) can be calculated as:

$$\text{Total Factor Productivity} = TFP = y/QI \tag{6.3}$$

Finally, consider a production in which multiple (m) outputs are produced by using multiple (n) inputs. In this situation, a total of $n \times m$ partial measurements of productivity can be calculated. It would, however, be more interesting to estimate a total measurement of productivity whereby all outputs are aggregated into an output index QO, and all inputs into an input index QI, and where the Total Factor Productivity (TFP) is then calculated as:

$$\text{Total Factor Productivity} = TFP = QO/QI \tag{6.4}$$

In the following, x and y are mainly considered scalars (one input and one output), but the results can be generalised to cover multiple inputs and multiple outputs, in which case x and y are interpreted as aggregates (input and output indices).

6.2.2 Efficiency

Efficiency can be briefly defined as the achieved compared to what can be achieved. If all the applied inputs could potentially produce 100 units, but only 80 units are produced, then the efficiency is 0.8, or 80 %. *Efficiency changes* means that the firm's position relative to the current technological frontier changes.

A production function $f(x)$ expresses per definition the maximum achievable output y when applying a given amount of input x. If the actual achieved quantity of output is called y_0 and the actual used quantity of input is called x_0, then efficiency is expressed as:

$$\text{Efficiency} = \frac{y_0}{f(x_0)} \tag{6.5}$$

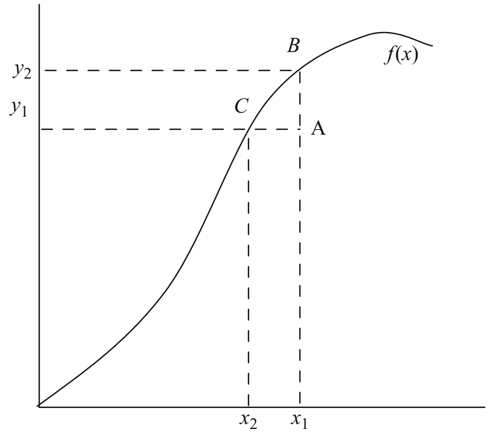


Fig. 6.2 Illustration of efficiency

The efficiency can be illustrated graphically, as shown in Fig. 6.2. The points B and C illustrate a production with an efficiency of 1 or 100 %. Points such as B and C are also sometimes referred to as *technical efficient* (Coelli et al. 2005). Point A, on the other hand, has an efficiency of less than 1 or less than 100 %. A production as illustrated by point A is also referred to as *technical inefficient*.

The degree of efficiency can be measured in two ways: One way is to measure it in the output dimension, i.e. express how much is produced compared to what could be produced. At point A, this would correspond to a measure expressed as the distance x_1A divided by the distance x_1B. Another way would be to measure the efficiency in the input dimension, i.e. to express how much input could be saved with the same produced output quantity. At point A, this would correspond to a measure expressed as the distance y_1C divided by the distance y_1A.

Efficiency can also be illustrated when there are two (or more) inputs. In Fig. 6.3, a production with two inputs has been illustrated. The points on the isoquant for the product amount y^0 are per definition an expression of a technically efficient production (the efficiency is 100 %), as it is not possible to produce more than y^0 with the given input combination. Point A (where an amount of precisely y^0 is produced) is, however, an expression of a technically inefficient production, as the same amount can be produced with less input. It is e.g. possible to produce the same amount in point C. The distance AC, or the distance 0 C divided by the distance $0A$, could be used as the efficiency measurement. However, please note that there are other ways of moving from point A to the isoquant than by going to point C. In practice, the inefficient producer should of course move to the point on the isoquant that is *economically efficient*, i.e. a point on the expansion path, which would depend on the price ratio. The ratio $0C/0A$ is often used in the literature as an expression for *technical efficiency*, as it is an entirely technical measure which can be established without knowledge of the economic (price) ratio (see more in Coelli et al. 2005).

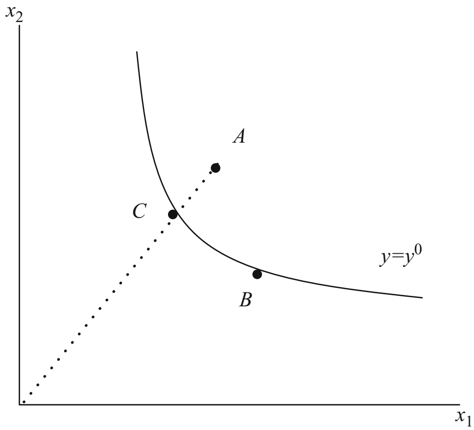


Fig. 6.3 Illustration of efficiency

6.2.3 Technological Changes

Technological (or *technical*) *change is defined as a shift in the production function over time,* or alternatively, technological change means that the frontier of the technology moves through time. If, at the point in time t, $y_t = f_t(x_t)$ and at a later point in time s, $y_s = f_s(x_s)$ and if $f_s(x^0) = \tau f_t(x^0)$, then the technological change (for the input amount x^0) over the period from t to s is defined as (τ-1) – or measured in percentages, $(\tau - 1) \times 100\%$.

Technological changes can be illustrated graphically, as shown in Fig. 6.4. As can be seen, the production function for the period s produces a higher yield than the production function for the previous period t for all input levels. For the input level x^0, this corresponds to a technological improvement of $(\tau - 1)$, where τ is y_{0s}/y_{0t}.

The technological changes can also be illustrated graphically when there are two inputs. Described in a figure with isoquants (see e.g. Fig. 6.3), the technological improvements could be illustrated by shifting the isoquant for a given output amount $y = y_0$ in the direction towards the zero point.

6.2.4 The Scale of Production

The scale of production identifies the point on the production function where production takes place. The essential issue in this context is whether production takes place in an area of the production function where there are *increasing returns to scale*, *decreasing returns to scale* or *constant returns to scale* (see also Fig. 4.4 in Chap. 4). The concepts are illustrated in Fig. 6.5.

The returns to scale at point c are increasing, as the production elasticity at this point is greater than one. As shown in Chap. 2 (Eq. 2.11), the production elasticity ε is calculated as:

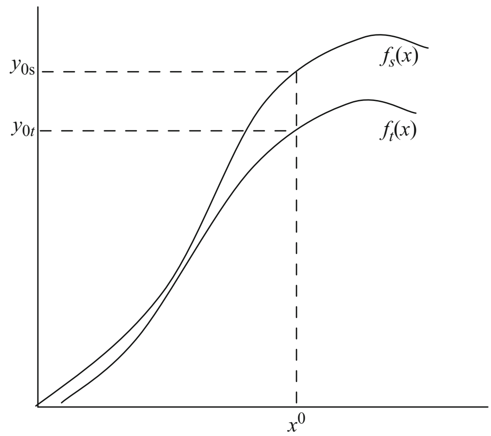


Fig. 6.4 Illustration of technological changes

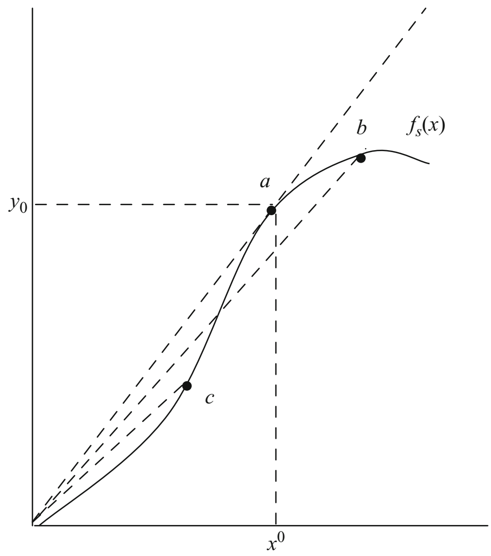


Fig. 6.5 Description of scale

$$\varepsilon = \frac{\frac{\partial y}{y}}{\frac{\partial x}{x}} = \frac{\partial y}{\partial x}\frac{x}{y} = \frac{MPP}{APP} \tag{6.6}$$

and the slope of the production function (*MPP*) around point *c* is greater than the slope of the line from the zero point (*APP*). The returns to scale around point *b* are, on the other hand, decreasing as the slope of the curve (*MPP*) here is less than the slope of the line from the zero point (*APP*). Finally, the returns to scale around point *a* are constant and equal to 1 as *MPP* here is equal to *APP*.[1]

The *highest productivity,* and thereby the highest output per unit of input, is achieved exactly at point *a*. Point *a,* or rather the input amount x_0, is therefore referred to as the *technically optimal scale of production*.

6.3 Changes in Productivity

With the already given definitions and descriptions, it is now possible to analyse and describe the reasons for productivity changes. The objective is to be able to explain and interpret changes in production of output and consumption of input, as these are the "raw data" that will be available to the practitioner/analyst in connection with the analysis of production-related relationships within an industry.

It should be noted that productivity changes themselves are not what is of most interest here. Rather it is the *reasons* for the productivity changes. Is an increasing production per input unit due to improved *efficiency*? Is it due to *technological improvements*? Or is it due to changes in the *scale of production*?

The point of reference is the original Fig. 6.1, and point *A* is assumed to describe the production and input consumption (according to the statistics) in the period *t*. It is, furthermore, assumed that the production in the subsequent period $t + 1$ is given by point *B* in Fig. 6.6 below. As can be seen, productivity has increased from A to B as the slope of a line through the zero point is larger for line 0*B* than line 0*A*. However, the question is; what is the reason for this? The three different possibilities are described in Figs. 6.7, 6.8, and 6.9.

In Fig. 6.7, the production function is assumed to be the same in period *t* and $t + 1$. The increase in productivity is therefore primarily due to improved efficiency. However, it should also be noted that the scale has changed so that both conditions have an influence.

In Fig. 6.8, the production is assumed to be efficient for both period *t* and period $t + 1$. The increase in productivity is primarily due to technical improvements. However, also here the change in scale has an influence.

In Fig. 6.9, the production is efficient both in period *t* and period $t + 1$. And there have been no technological changes. The change in productivity is due exclusively to a change in scale and, in this example the producer uses a (technical) optimal scale in period $t + 1$.

The described division of the changes in productivity into the three components, as illustrated graphically in Figs. 6.7, 6.8, and 6.9, can be derived mathematically (Coelli et al. 2005).

[1] Please note that the concept of *returns to scale* is formally associated with a description of what happens to the production when *all* inputs are increased by a certain factor. Hence, in the example here, *all inputs* consist of only one input.

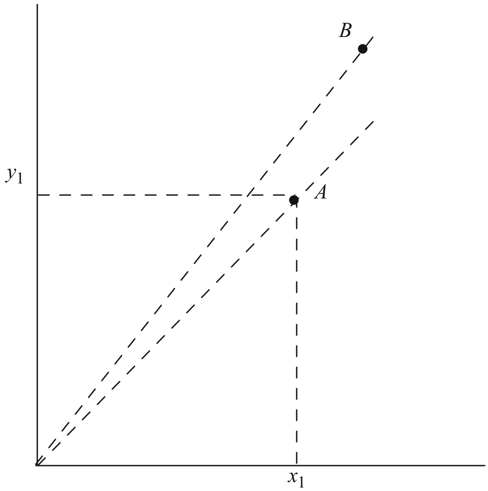


Fig. 6.6 Productivity increase

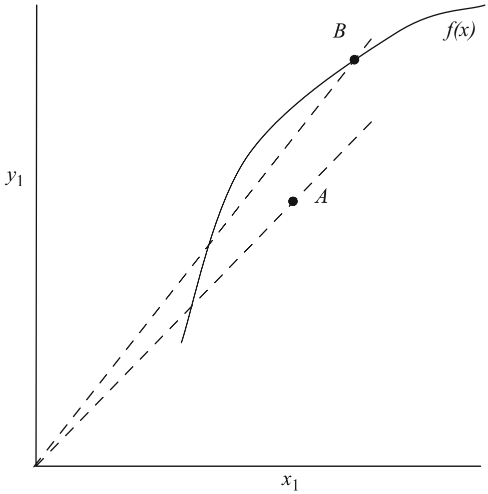


Fig. 6.7 Efficiency increase

The productivity in period t, when the productivity in period s is set equal to 1, is:

$$P_{st} = \frac{y_t/x_t}{y_s/x_s} \tag{6.7}$$

The actual measured output y_t can be expressed as:

$$y_t = \tau_t f_t(x_t) \tag{6.8}$$

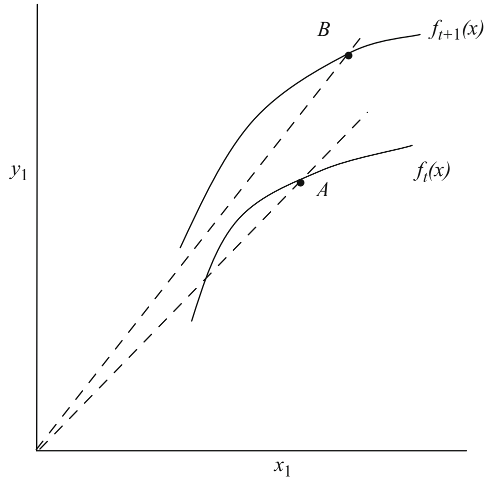


Fig. 6.8 Technological change

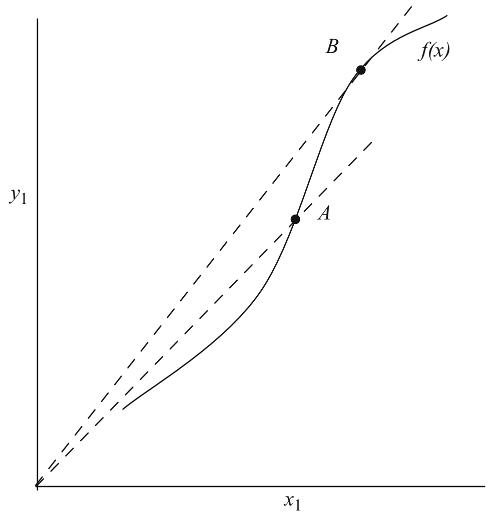


Fig. 6.9 Change of scale

in which τ_t is the expression for the efficiency in period t. The same is true for period s. Inserting Eq. 6.8 in Eq. 6.7 yields:

$$P_{st} = \frac{\tau_t}{\tau_s} \times \frac{f_t(x_t)/x_t}{f_s(x_s)/x_s} \tag{6.9}$$

If the consumption of x is the same for both periods ($x_t = x_s = x^0$), the productivity shown in Eq. 6.9 can be decomposed into the following two factors:

$$P_{st} = \frac{\tau_t}{\tau_s} \times \frac{f_t(x^0)}{f_s(x^0)} \tag{6.10}$$

in which the first fraction measures the change in efficiency, and the second fraction measures the technical change at the input level x^0.

Equation 6.10 can be expanded to accommodate different input consumption (input scale) in period t and period s. If we look at only one input (or input vectors where all inputs are changed by the same factor), the relation between input in the two periods can be written as $x_t = \kappa x_s$, where κ is a positive number. If the input consumption in period t is higher (which is presupposed here), κ is greater than 1. It is furthermore presupposed that the production function $f_t(x)$ is homogeneous of degree ε at the input level x_t. Hence, Eq. 6.9 can be written as:

$$P_{st} = \frac{\tau_t}{\tau_s} \times \frac{f_t(\kappa x_s)/\kappa x_s}{f_s(x_s)/x_s} = \frac{\tau_t}{\tau_s} \times \kappa^{\varepsilon-1} \times \frac{f_t(x_s)}{f_s(x_s)} \tag{6.11}$$

because functions that are homogeneous of degree ε can be written as:

$$\frac{f_t(\kappa x_s)}{\kappa x_s} = \kappa^{\varepsilon} \times \frac{f_t(x_s)}{\kappa x_s}$$

In addition to the two components, changes in efficiency (the first component in Eq. 6.11), and technological changes (the last component in the right hand side of Eq. 6.11), there is one additional component $\kappa^{(\varepsilon-1)}$ expressing the scale effect, as illustrated in Eq. 6.11. If the production function is homogeneous of degree one ($\varepsilon = 1$) locally (i.e. for the observed input–output combinations), then the factor $\kappa^{(\varepsilon-1)}$ is equal to 1, and the changes in scale do not affect productivity. Hence, in such cases, the changes in productivity are solely due to changes in efficiency and changes in technology.

References

Balk, B. M. (1998). *Industrial price, quantity, and productivity indices*. Boston: Kluwer Academic Publishers

Coelli, T., Prasada Rao, D. S., O'Donnell, C. J., & Battese, G. (2005). *An introduction to efficiency and productivity analysis* (2nd ed.). New York: Springer

Input Demand Functions

7

7.1 Introduction

In this chapter, the theory introduced in Chap. 4 will be used to derive the company's demand for input used in production. Furthermore, how the theory can be used to analyse what happens to the demand for input when the relative prices vary will also be examined.

The representation in the first sections of this chapter presupposes a market with perfect competition, i.e. the company is a price taker and does not have the possibility of influencing the market price for the required inputs. At the end of the chapter (Sect. 7.5), the input demand under the more general assumption that the price for input can vary, depending on the amount demanded by the company, is discussed.

What conditions determine how much of the variable input factor x_1 a company will buy and use?

First of all, the price (w_1) must be a key factor. The higher the price, the smaller the amount the company will be expected to buy. However, this will probably depend on whether it is possible to use other (cheaper) inputs instead. Hence, the price for other variable inputs ($w_2, \ldots, w_k$) must also be a key factor. Furthermore, there will be the price of output (p_y); the higher the price of output, the more input the company will be expected to acquire. However, this will probably depend on how the production function appears – i.e. how much more output would be generated by adding more input. Therefore, the form of the production function (parameters (α)) will have an influence. Finally, one could imagine that there are budget constraints, so that it is not possible to buy the entire quantity that is generally required. Hence, a budget constraint (C^0) can be a key factor. In conclusion, the point of reference regarding the input demand function is expected to be a function with the following parameters:

$$x_1 = x_1(p_y, w_1, \ldots, w_k, \alpha, C^0) \tag{7.1}$$

S. Rasmussen, *Production Economics*, Springer Texts in Business and Economics,
DOI 10.1007/978-3-642-30200-8_7,

How such a functional relationship can be derived is demonstrated in the following.

7.2 One Variable Input

The point of reference for the analysis is the criterion for profit maximisation derived in Chap. 3 (see Eq. 3.2):

$$\frac{p_y MPP_1}{w_1} = \frac{VMP_1}{w_1} = 1 \tag{7.2}$$

where VMP_1 (the value of the marginal product for input 1) is the marginal product MPP_1 multiplied by the product price p_y, and where w_1 is the price of input 1. The criterion means that the optimal level of input is where $VMP_1 = w_1$ at the decreasing part of the VMP-curve (see Sects. 3.1 and 3.2 in Chap. 3).

The criterion is illustrated graphically in Fig. 7.1, in which the initial price w_1 is presumed to be *equal to* w_{10}, and the optimal application of input x_1 therefore equals x_{10}.

If the price falls to w_{11}, then the optimal application of (and thereby the demand for) x_1 increases to x_{11}. If the price increases to w_{12}, then the optimal application of (and thereby the demand for) x_1 falls to x_{12}.

It follows that if the producer maximises profit, then the VMP curve represents the relationship between the input price and the corresponding demand for the same input. *Hence, the VMP curve is identical to the demand curve for input.*

As the marginal product MPP depends on x and the parameters (α) of the production function, the criterion $VMP_1 = w_1$ can be written $p_y MPP_1(x_1, \alpha) = w_1$. Solving for x_1 using the implicit function theorem provides the solution:

$$x_1 = x_1(p_y, w_1, \alpha) \tag{7.3}$$

Hence, the demand for a variable input x_1 is a function of the price of output, the price of the input in question, and the production function parameters. Please note that Eq. 7.3 is based on the precondition that x_1 is the only variable input (all other inputs are presumed to be fixed) and that there are no budget constraints.

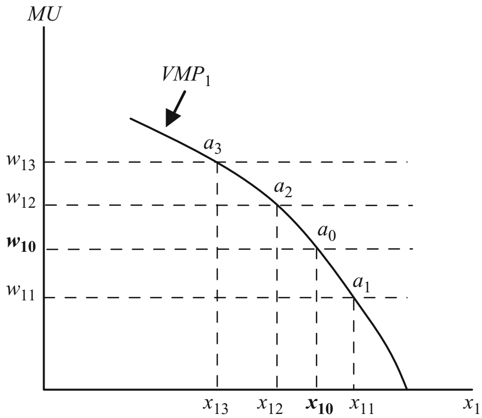
Example 7.1

We use a Cobb-Douglas production function:

$$Y = f(x_1) = A{x_1}^b \tag{7.4}$$

The production function vector of parameter α is (A, b). The marginal product is:

Fig. 7.1 Demand function for input



$$MPP = bAx_1{}^{(b-1)} \tag{7.5}$$

and the criterion for profit maximisation is therefore:

$$p_y bAx_1{}^{(b-1)} = w_1 \tag{7.6}$$

Isolating x_1 generates:

$$x_1 = w_1^{(1/(b-1))} p_y^{(-1/(b-1))} (bA)^{(-1/(b-1))} \tag{7.7}$$

If the parameter values (the vector α) e.g. are given the values $A = 1$ and $b = 0.5$ and inserted in Eq. 7.7, the *input demand function* can be expressed as:

$$x_1 = 0.25 p_y^2 / w_1^2 \tag{7.8}$$

which, for a given output price, is a decreasing function in w_1 and, for a given input price, is an increasing function in p_y. Hence, the demand for input decreases (increases) with an increasing (decreasing) input price and increases (decreases) with an increasing (decreasing) output price.

The demand for input can also be expressed by *the demand elasticity*. *The demand elasticity ε_D is defined* as the relative (percentage) change in the demand at a relative (percentage) change in the input price, or formally:

$$\varepsilon_{D1} = \frac{dx_1/x_1}{dw_1/w_1} = \frac{dx_1}{dw_1}\frac{w_1}{x_1} = \frac{d\ln x_1}{d\ln w_1} \tag{7.9}$$

The expression illustrated is referred to as *the own-price elasticity* which is the (relative) change in the demanded quantity, when the price being changed is the *own-price* of the input in question (as opposed to *the cross-price elasticity* where it is the price of another input that is changed (discussed later)).

Example 7.2

Consider the previous example (see Eq. 7.7) and calculate the own-price elasticity using the middle term formula in Eq. 7.9. Differentiate first Eq. 7.7 with regard to w_1, which produces:

$$\frac{dx_1}{dw_1} = \frac{1}{b-1}\frac{x_1}{w_1} \tag{7.10}$$

This is then multiplied by w_1/x_1, and the resulting own-price elasticity is therefore equal to:

$$\varepsilon_{D1} = \frac{dx_1}{dw_1}\frac{w_1}{x_1} = \frac{1}{b-1} \tag{7.11}$$

The own-price elasticity could also be calculated using the last formula element in Eq. 7.9. Taking the logarithm of x_1 in Eq. 7.7 gives:

$$\ln x_1 = \frac{1}{b-1}\ln w_1 - \frac{1}{b-1}\ln p_y - \frac{1}{b-1}(\ln b + \ln A) \tag{7.12}$$

and differentiating it with regard to ln w_1 gives:

$$\varepsilon_{D1} = \frac{d\ln x_1}{d\ln w_1} = \frac{1}{b-1} \tag{7.13}$$

which is the easiest way to calculate the own-price elasticity.

If the above parameter values ($b = 0.5$) are used, you will find that the result here is an own-price elasticity of -2. If the input price is increased by 10 %, the demand will therefore fall by 20 %.

You can also calculate the output-price elasticity. *The output-price elasticity* ε_{Dy} is the relative change in the demand for an input when the price of output is changed and calculated in accordance with formula Eq. 7.9 as:

$$\varepsilon_{Dy} = \frac{d\ln x_1}{d\ln p_y} \tag{7.14}$$

Differentiating Eq. 7.12 gives:

$$\varepsilon_{Dy} = \frac{d\ln x_1}{d\ln p_y} = -\frac{1}{b-1} \tag{7.15}$$

A parameter value of $b = 0.5$ gives an output-price elasticity of 2. If the output price increases by e.g. 10 %, then the demand for input is increased by 20 %.

7.3 Multiple Variable Inputs

When using multiple variable inputs the price changes for an input may not only affect the demand for the input in question but also the demand for other inputs. Using multiple variable inputs, it is possible to adjust the production so that input that has experienced a price increase can be replaced by input that has now become comparatively cheaper.

This substitution has previously been illustrated in Chap. 4 in which Fig. 4.2 shows that the optimal combination of two variable inputs depends on the relative prices, and that an increased price for input x_1 entails that a given amount of y can be produced by using more of x_2 and less of x_1. Hence, price changes will entail that the producer will adjust the production along the isoquant – i.e. cut down on the input that is experiencing a price increase.

However, changes in the relative prices also have other implications. Moving along the isoquant will also produce changes in the marginal products (*MPP*). This means that the production y, which previously entailed a profit maximum, now has to be adjusted as the profit maximum is to be found on another isoquant. However, whether this is the case depends on whether the changes in the consumption of one input affect the marginal product of other inputs.

In an effort to describe the interaction between various inputs, the two inputs i and j can be described as being *complementary*, *competitive,* or *independent*. The definition is as follows:

Complementary inputs: $\frac{\partial MPP_i}{\partial x_j} > 0$

Competitive inputs: $\frac{\partial MPP_i}{\partial x_j} < 0$

Independent inputs: $\frac{\partial MPP_i}{\partial x_j} = 0$

An example of complementary inputs could be e.g. water and nitrogen fertiliser for growing crops. Here the effect of the fertiliser is improved by the irrigation of dry land. Another example is labour and management, whereby the productivity of labour is improved by increasing the amount of management. An example of competing inputs are inputs, which could very easily replace each other – for example, nitrogen in the two nitrogen fertilisers, nitrate and ammonia. Another example is fuel and electricity, both used for the heating of buildings. It is up to the reader to find examples of independent inputs and also to find further examples of complementary and competitive inputs.

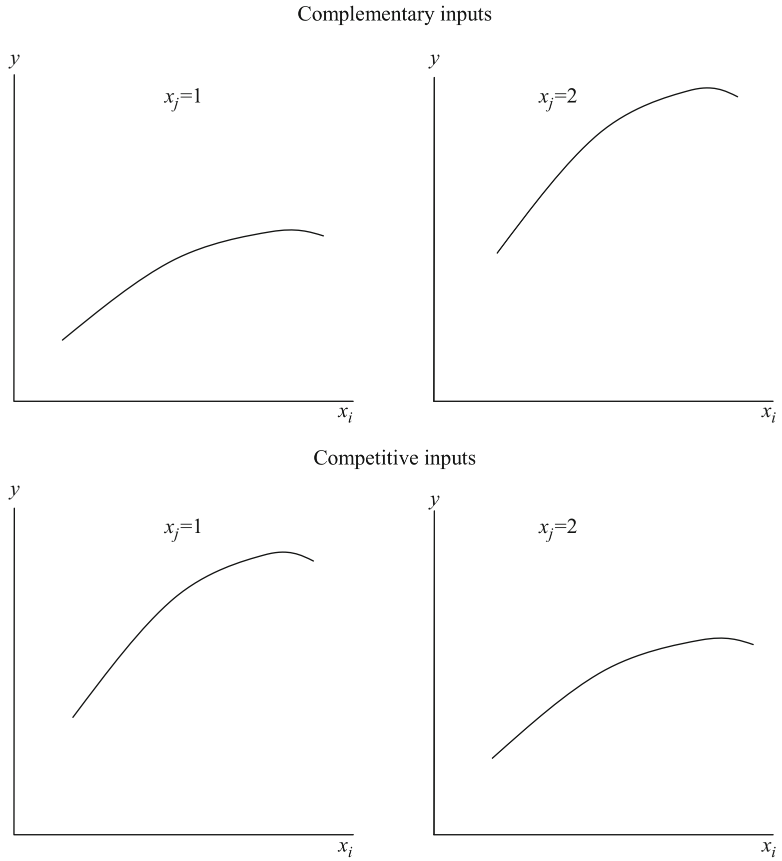


Fig. 7.2 Interaction between input i and input j

As illustrated in the top part of Fig. 7.2, the production function $y = f(x_i|x_j)$ has a larger slope (MPP_i) at an increased supply of x_j. Hence, the two inputs "support" each other – are complementary. In the lower part of the figure, an increased amount of x_j results in the production function $y = f(x_i|x_j)$ becoming flatter, i.e. MPP_i decreases as x_j is increased – the two inputs are competitive.

For independent inputs, the production function $y = f(x_i|x_j)$ is independent of the amount of x_j.

Graphically, this relationship can be illustrated as shown in Fig. 7.2.

The derivation of the demand function for an input, when there are multiple inputs, is – as before – based on the criterion for profit maximisation derived in Chap. 4. The following criteria are true for profit maximisation (see Eqs. 4.18a and 4.18b) for two variable inputs:

$$w_1 = MPP_1 p_y (\equiv VMP_1) \tag{7.16a}$$

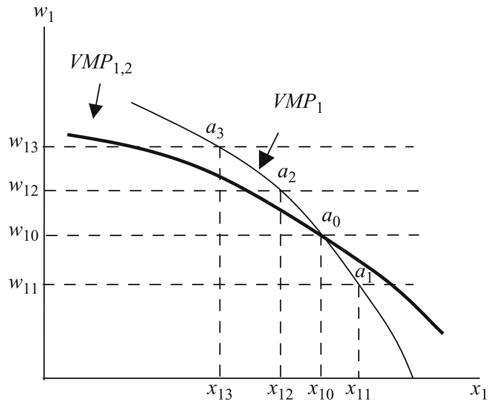


Fig. 7.3 Demand function for input x_1 with two variable inputs

$$w_2 = MPP_2 p_y (\equiv VMP_2) \tag{7.16b}$$

For each of the two inputs, the criterion can be illustrated graphically as previously shown in Fig. 7.1. With two variable inputs, there will, however, be a simultaneous adjustment of both inputs in connection with profit maximisation, so that any substitution between the two inputs will have an influence. For "normal" inputs, this means that the effect of a price change would be larger with multiple variable inputs, as price increases will in fact give rise to a substitution of some of the now more expensive inputs with other inputs. The effect of price changes is illustrated in Fig. 7.3, which is similar to Fig. 7.1, with the sole difference that the *VMP* curve for input x_1 has a flatter shape ($VMP_{1,\,2}$) when there are two variable inputs.

Hence, the VMP curve is identical to the demand curve for input. However, the shape of the VMP curve depends on which of the other inputs are considered to be variable. The better the possibility for substitution, the greater the effect of the price change on a given input (the *VMP* curve in Fig. 7.3 turns counter clockwise).

Let's have a look at the factors that this demand is dependent on. As can be seen from the criterion (7.16a and 7.16b), both the output price p_y as well as the input prices w_1 and w_2 are part of this relationship. Add to this the marginal products MPP_1 and MPP_2 which both contain the production function parameters (α). The demand function can therefore be expressed as in the following general function expression:

$$x_1 = x_1(p_y, w_1, w_2, \alpha) \tag{7.17}$$

Hence, the demand for a variable input x_1 is a function of the price of output, the price of the input in question, the price of other variable inputs (here w_2), and the production function parameters. Please note that Eq. 7.17 is based on the precondition that there are no budget constraints.

Example 7.3

The use of a Cobb-Douglas production function is presupposed:

$$y = Ax_1^a x_2^b \tag{7.18}$$

The production function vector of parameters, α, is (A, a, and b). The criterion for profit maximisation Eq. 7.16a is:

$$ap_y A x_1^{a-1} x_2^b = w_1 \tag{7.19a}$$

and Eq. 7.16b:

$$bp_y A x_1^a x_2^{b-1} = w_2 \tag{7.19b}$$

Isolating x_1 in Eq. 7.19a generates:

$$x_1 = w_1^{(1/(a-1))} (ap_y A)^{-1/(a-1)} x_2^{(-b/(a-1))} \tag{7.20}$$

from which it appears that the demand for input x_1 depends on the input price w_1, the output price p_y, the amount of other variable inputs x_2, and the production function parameters a, b, and A.

The problem with the demand function for input x_1 in Eq. 7.20 is that the amount of the other variable inputs x_2 also depends on the price w_1 of input x_1. It is therefore not possible to differentiate Eq. 7.20 with regard to w_1 before the functional relationship between x_2 and w_1 is established.

The method for this is a simultaneous solution of Eqs. 7.19a and 7.19b. Dividing Eq. 7.19a by Eq. 7.19b generates:

$$x_2 = w_1 \frac{bx_1}{aw_2} \tag{7.21}$$

which is the expression of the expansion path. Inserting this expression of x_2 in Eq. 7.20 generates the following demand function for input x_1:

$$x_1 = w_1{}^{(1-b)/(a+b-1)} w_2^{b/(a+b-1)} (p_y A)^{-1/(a+b-1)} a^{(b-1)/(a+b-1)} b^{-b/(a+b-1)} \tag{7.22}$$

Taking the logarithm and differentiating with respect to the logarithm yields:

$$\varepsilon_{D1} = \frac{d \ln x_1}{d \ln w_1} = \frac{(1-b)}{(a+b-1)} \tag{7.23}$$

which is less than zero when $a + b$ is less than 1, i.e. when the returns to scale are decreasing (see Chap. 4). Hence, the own-price elasticity for input, when the

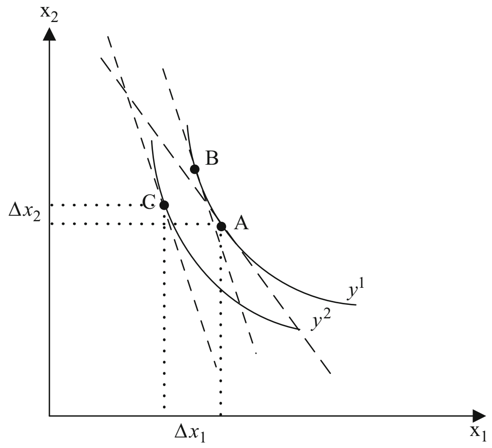


Fig. 7.4 Substitution effect and income effect

production function is a Cobb-Douglas function, is negative when the returns to scale is decreasing.

Comparing Eq. 7.23 with Eq. 7.13 shows that the demand elasticity for input x_1 depends on whether there are other variable inputs. As mentioned before, the effect would normally be greater when there are other variable inputs.

This relationship can be illustrated graphically, as shown in Fig. 7.4. The initial price ratio corresponds to the dotted line with the tangent point at point A. The price w_1 of input x_1 increases and the new price ratio is given by the dotted line through point B (or C). The profit maximum is, initially, presumed to be achieved through the production of the product amount y^1. Furthermore, it is presumed that it is optimal to produce the amount y^2 (point C) after the increase of the price of input x_1.

As can be seen, the price increase first results in a *substitution*, so that less x_1 and more x_2 is used for a given production (movement along the isoquant from A to B (*substitution effect*)). However, point B is not optimal as the high price level implies that it is now no longer profitable to produce the amount y^1. The supply of both x_1 and x_2 is reduced, and the final production after the adaptation to the new price ratios is y^2 in point C (from B to C, *income effect*).

The total adjustment described here entails that the consumption of x_1 decreases and the consumption of x_2 increases, which is an indication of substitution between the two inputs.

Generally speaking, the demand for an input decreases when the price of the input increases. However, it is *not* possible to draw any general conclusions about the effect of the use of *other variable inputs*. In the graphical example in Fig. 7.4, the consumption of input x_2 increases when the price of input x_1 increases. This might not always be the case though. There can be situations where the increase in the price of an input not only results in a decrease in the amount of the input in question but also a decrease in the amount of other inputs.

The change in the demand for an input when the price of another input is changed is referred to as *the cross-price elasticity*. The cross-price elasticity between input i and input j is defined by:

$$\varepsilon_{Dij} = \frac{\frac{dx_j}{x_j}}{\frac{dw_i}{w_i}} = \frac{dx_j}{dw_i}\frac{w_i}{x_j} = \frac{d\ln x_j}{d\ln w_i} \tag{7.24}$$

It is not possible to say something general about the sign of this expression.

7.3.1 Increasing Output Price

What happens with the consumption of input when *the output price p_y increases*?

When the output price increases the producer will – everything else being equal – increase the production of y (see Fig. 5.4). And an increased production of y presupposes the use of more input.

Normally, increasing production of y will be a result in increasing consumption of *all* inputs. However, this might not always be the case. It is possible to have production conditions where increasing production of y results in decreasing use of one or more inputs.[1] In Fig. 7.4, you will find an example where an increase in the production from y^2 (point C) to y^1 (point A) entails that the consumption of x_2 in fact *decreases* (however, the consumption of x_1 increases in return).

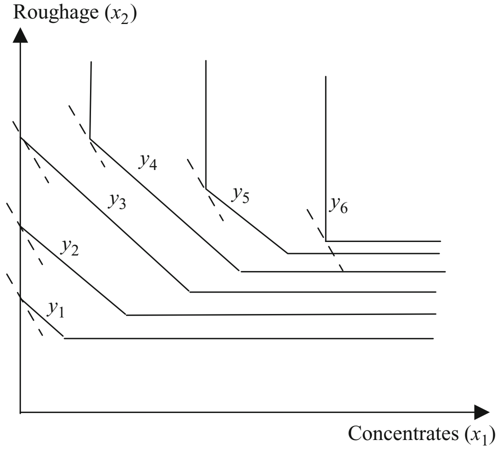
Inputs, the consumption of which increases when production increases, are called *normal inputs*, whilst inputs, the consumption of which decreases when production increases, are called *inferior inputs*. In Fig. 7.4, both x_1 and x_2 are normal inputs as the consumption of both inputs increases when production increases (at given input prices).

In practice, there are not that many examples of inferior inputs. However, the production of milk with the use of two kinds of fodder "roughage" and "concentrates" is a relevant example within farming (the example is borrowed from Flaten, 2001).

The example is illustrated in Fig. 7.5 in which the isoquants are drawn as piecewise linear curves. Within certain intervals, roughage and concentrates can basically replace each other in the ratio 1:1 (sloping part of isoquants). However, due to biological conditions, this substitution is only possible within limited intervals. Eventually, the isoquants become vertical/horizontal. Milk production can be increased from y^1 to y^6 by increasing the amount of fodder (roughage or concentrates). The price ratio between roughage and concentrates is illustrated by the dotted lines. At low milk production (y^1–y^3), the milk can be produced solely by

[1] The consumption of one or more of the inputs must, however, necessarily increase for the production of y to increase. In situations with two (variable) inputs, the consumption of one of the inputs will thus always increase.

Fig. 7.5 Isoquants for milk production



the application of roughage (e.g. grass). However, if the amount of milk is to be increased to more than y^3, part of the fodder should be in the form of more easily digestible and energy rich concentrates. To allow room for absorption of increasing amounts of concentrate the supply of roughage must be reduced, and at the production level y^6 the use of roughage is reduced considerably while the application of concentrates is increased heavily. Hence, after reaching a certain level, an increasing production will result in a decreasing roughage application, and roughage will thus be an inferior input here.

7.4 Input Demand Under Budget Constraint

In the above, the company was assumed to have the possibility of buying inputs without constraints. However, sometimes, there may be budget constraints and the question is then how this affects the adaptation when the input price increases.

The conditions are outlined in Fig. 7.6. The budget constraint is C^0 and the initial budget line is given by the flattest of the two budget lines through C^0/w_2. The price of input x_1 is now assumed to increase so that the budget line is given by the steeper line through C^0/w_2 after the price increase.

In the first situation (A), the demand for both input x_1 and input x_2 decreases. In the second situation (B), the demand for input x_1 decreases, while the demand for input x_2 increases. In the last situation (C), the demand for input x_1 decreases, while the demand for input x_2 is unchanged.

Hence, it appears that while the price increase of one input will always result in a lower demand for the input in question, the effect on other variable inputs will be higher, lower, or unchanged demand.

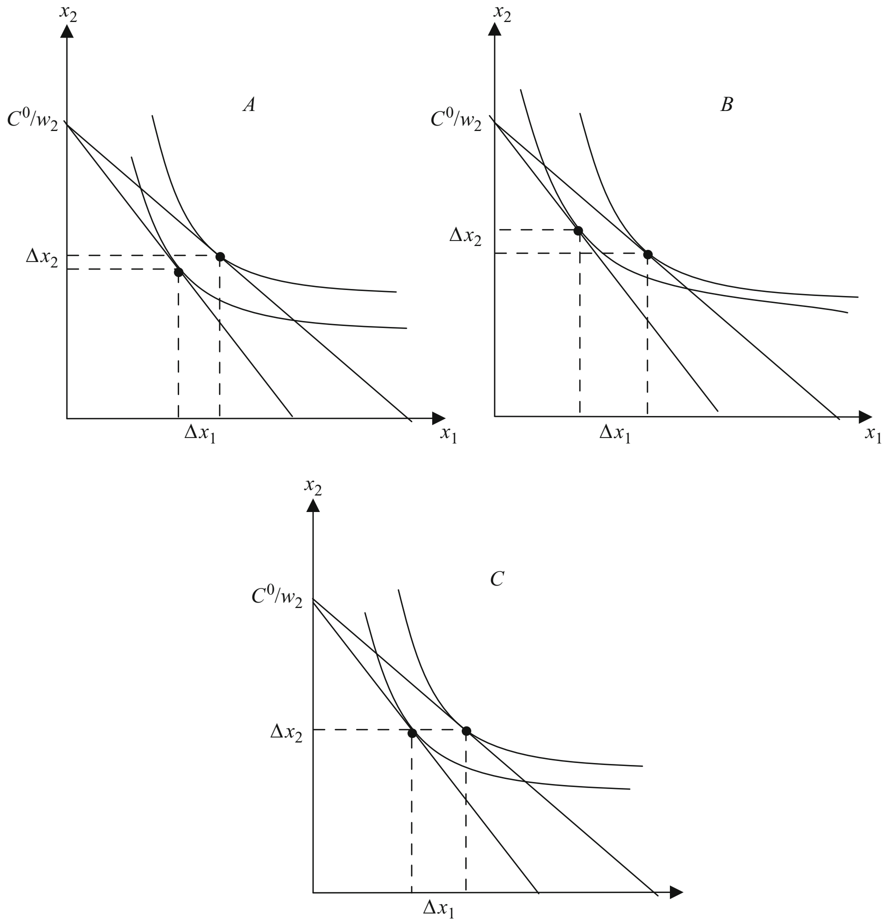


Fig. 7.6 Substitution under budget constraint

7.5 Demand When the Input Price Depends on the Demand

In this last section, the precondition for perfect competition on the factor market is abandoned, as each individual company is now presupposed to be big enough for its demand to affect the input price. (We still assume that the output market is competitive, i.e. the producer is a price taker on the output market). In principle there are two possibilities: The input price w increases with the increasing demand for x, i.e. $\mathrm{d}w/\mathrm{d}x > 0$. The other possibility is that the input price decreases with the increasing demand, i.e. $\mathrm{d}w/\mathrm{d}x < 0$. This last possibility is e.g. relevant when a company, due to its size, is eligible for a quantity discount in connection with bulk buying.

The demand function can be derived based on the expansion path, which is derived as before, using the Lagrange function L which is maximised with respect to the two variable inputs x_1 and x_2 (compare with Eqs. 4.10, 4.11, and 4.12 in Chap. 4):

$$L = f(x_1, x_2) + \theta(C - (w_1(x_1)x_1 + w_2(x_2)x_2) \tag{7.25}$$

We use the term $w(x)$ to indicate that the input price w is a function of x. The maximisation with regard to the two variables x_1 and x_2, as well as the Lagrange multiplier θ, is done by taking the partial derivatives and setting them equal to zero. This produces the following three conditions for an optimal solution:

$$MFC_1 = w_1(x_1) + \frac{\partial w_1}{\partial x_1}x_1 = MPP_1/\theta \tag{7.26a}$$

$$MFC_2 = w_2(x_2) + \frac{\partial w_2}{\partial x_2}x_2 = MPP_2/\theta \tag{7.26b}$$

$$C = (w_1(x_1)x_1 + w_2(x_2)x_2 \tag{7.26c}$$

Dividing Eq. 7.26a by Eq. 7.26b produces the necessary condition for the maximisation of Eq. 7.25 for the given C:

$$\frac{MFC_1}{MFC_2} = \frac{MPP_1}{MPP_2} \tag{7.27}$$

where MFC_i stands for *the marginal factor costs* for the input i calculated as the intermediate expression after the first equal sign in Eqs. 7.26a and 7.26b. The marginal factor costs are expressed as the incremental cost in connection with the purchase of one more unit of input. The marginal factor costs can also be expressed as:

$$MFC_i = w_i(x_i)(1 + E_{x_i}) \tag{7.28}$$

in which E_{x_i} is the price elasticity for input x_i given by:

$$E_{x_i} \equiv \frac{\frac{\partial w_i}{w_i}}{\frac{\partial x_i}{x_i}} = \frac{\partial w_i}{\partial x_i}\frac{x_i}{w_i} \tag{7.29}$$

If the price elasticity for an input is zero, then the marginal factor cost in Eq. 7.28 is equal to the factor price w_i, corresponding to perfect competition. If the price elasticity is positive, the price the company owner pays increases with the increase in purchased input. However, if the price elasticity is negative, it is possible to achieve a lower price with an increasing amount.

The criterion for profit maximisation is generalised similarly when the possibility of varying input prices is included. As in formula (4.19) in Chap. 4, the criterion for profit maximisation under varying input prices is thus equal to:

$$\frac{VMP_1}{MFC_1} = \frac{VMP_2}{MFC_2} = \ldots = \frac{VMP_n}{MFC_n} = 1 \tag{7.30}$$

where MFC_i is given in Eq. 7.28.

The case of non-competitive output markets are treated in Chap. 13.

Reference

Flaten, O. (2001). *Økonomiske analyser av tilpasninger i norsk mjølkeproduktion*. Dr. Scient Thesis from Institut for økonomi og samfunnsfag, Norges landbrukshøgskole, Ås

8 Land and Other Inputs

8.1 Introduction

In the previous chapters we derived conditions for a cost minimising combination of inputs (Chap. 4), and studied how the demand for variable input depends – not only on the input price, but also on the prices of other variable inputs that may be used to substitute the input in question (Chap. 7). However, there are special cases/inputs when it is not possible to apply the previous models directly, and the concept of the pseudo scale line becomes useful.

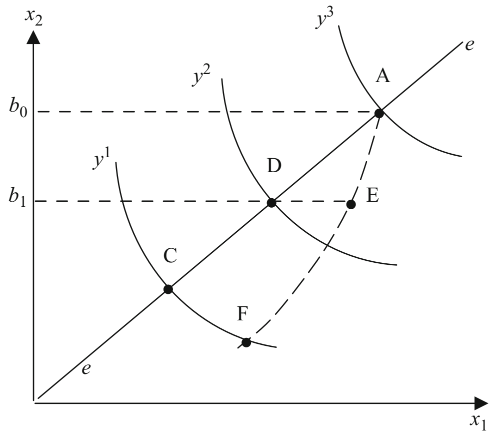
8.2 Land as a Special Input

Land is a special input. It is special in the sense that it is always acquired and thus available in a certain amount *before* the other inputs are added. For instance, when the farmer grows wheat, he *first* buys (or rents) land *and then* he applies the seed, fertilisers, etc. that are necessary inputs to grow wheat. A car manufacturer in Sweden with its cold climate *first* builds and insulates the factory buildings, *and then* he decides how much fuel to buy and use for heating the buildings during the production process. Thus, even though the amount of land, seed and fertiliser may all be variable inputs, the optimal combination of land and seed or land and fertiliser is *not* determined according to the principle in Eq. 4.19 in Chap. 4. And the optimal combination for the insulation of the car factory building and fuel for heating is *not* determined according to the principle in Eq. 4.19 in Chap. 4.

To see why, consider land as an input in agricultural production. Besides land (x_2), consider for simplicity's sake that there is only one other input (x_1), which is an aggregate of all the other inputs except land. In the long run, when the farmer has the possibility of buying and selling land, both land (x_2) and "other input" (x_1) are variable inputs.

S. Rasmussen, *Production Economics*, Springer Texts in Business and Economics,
DOI 10.1007/978-3-642-30200-8_8,

Fig. 8.1 Isoquants and pseudo-scale line



Assume that this long run adjustment has taken place in the "normal way" as described in Chap. 4, i.e. along the expansion path *ee* in Fig. 8.1.[1] The optimal combination of "other input" and land expanding the production along the expansion path ee is the point A in Fig. 8.1 where $VMP_1 = w_1$ and $VMP_2 = w_2$.

However, land cannot be combined with other inputs used for cultivating land (i.e. fertilisers, pesticides, irrigation etc.) in the same way that variable input would normally be combined. In the example presented here, these "other inputs" are added to the land, and land must therefore – per definition – be present as a fixed factor when it is decided how much of the "other input" should be added. Just think of the application of fertiliser when the acreage is given at the time when the decision is made as to how much fertiliser should be applied. With land as the fixed input in relation to "other input", the adjustment of the amount of land and "other input" therefore takes place along the pseudo scale line, which is the dotted curve AEF, as illustrated in Fig. 8.1, and *not* along the expansion path *ee*.

To supplement the graphical representation in Fig. 8.1 above, consider the following mathematical representation. The production is described by a production function:

$$y = f(x_1, x_2) \tag{8.1}$$

where y is the yield (e.g. kg of cereal crops), x_1 is "other input" (the aggregate of fertilisers, pesticides, labour, machinery etc.), and x_2 is hectares of land.

With regard to the subsequent analysis, there is no problem in combining all other inputs but land into one aggregate input, x_1, which we briefly describe as the

[1] For the sake of simplicity, the expansion path ee is drawn as a straight line (compare Fig. 4.5 in Chap. 4).

"other input". However, it may be helpful to stop and consider the underlying assumptions when you make such a simplification.

Firstly, the "other input" (x_1) is now an *aggregate of a number of inputs*. It represents a sum of all these inputs such as fertilisers, pesticides, seeds, labour, machinery capacity etc. But what are the units in which x_1 is actually measured? One can hardly just add up kg, litres, hours etc. and use this as the input measure.

No, this is not what one would normally do. Instead, x_1 should be calculated as a quantity *index* where the general form of the calculation of a quantity index Q is:

$$x_1 = Q = Q(x_{11}, \ldots, x_{1p}) \tag{8.2}$$

Here, the function Q is the function used for aggregating all the p inputs (fertilisers, pesticides, etc.) that are parts of x_1.

A brief introduction to quantity indices was included in Chap. 6 (see Sect. 6.2.1), where the formula for the calculation of a so-called Laspeyres quantity index was introduced. Further issues in connection with the calculating of the relevant quantity indices are not discussed in further detail here. Please refer to the extensive literature about index theory (see e.g. Balk 1998). It should however be mentioned that the function Q can in fact be interpreted as a production function which, based on all the p inputs $x_{11}, \ldots, x_{1p}$, "produces" the (intermediate) "product" x_1 which is then used as an input in the final production function f in Eq. 8.1. Hence, it is possible to say that the function Q "produces" the basket of input x_1 which is then used for the final production (of for instance cereal crops).

The precondition for the use of an input aggregate (an index) as an independent input in a production function as f in Eq. 8.1 is that there is a certain degree of independence between the inputs that are part of the index x_1 and the other input, land (x_2). This independence requirement can be formally formulated as:

$$\frac{\partial\left(\frac{MPP_{1i}}{MPP_{1j}}\right)}{\partial x_2} = \frac{\partial MRS_{ij}}{\partial x_2} = 0 \qquad (\text{for all } i \text{ and } j) \tag{8.3}$$

The condition (8.3) implies that the actual production technology should be of such a nature that the marginal rate of substitution (MRS) between any two inputs of the inputs being aggregated is independent of the amount applied of the other input (x_2). In this present example, the condition thus entails that the marginal rate of substitution (the slope of the isoquant) between e.g. fertilisers and pesticides should be independent of the amount of land used as input (see also Chambers (1988), Chap.5).

Compared to practice, this precondition is hardly unreasonable. However, the reader is encouraged to assess whether there are observations which are not consistent with this precondition in practice.

8.3 Example of Homogeneous Production Function

After this small digression, we will now return to the mathematical representation of the function f in Eq. 8.1. Assume that the production function f is homogeneous of degree one. In this case, the production function f can be expressed as:

$$f(tx_1, tx_2) = tf(x_1, x_2) \tag{8.4}$$

cf. the discussion in Sect. 4.3. The assumption that f is homogeneous of degree one can hardly be said to be entirely unreasonable in this example. In reality, this entails that each time the acreage is expanded by one hectare, and the same amount of the "other input" (x_1) as for all previous hectares is added to this extra hectare, then the total yield y is increased by an amount corresponding to the average yield of the previous hectares. The assumption of homogeneity is not decisive but facilitates an easier representation in the following.

As mentioned before, the optimal supply of x_1 is determined *after* the amount of x_2 has been chosen. Therefore x_2 is a fixed input (and thus a constant), and t can therefore be set equal to $1/x_2$ in Eq. 8.4, which means that Eq. 8.4 can be expressed as:

$$z = \frac{y}{x_2} = f\left(\frac{x_1}{x_2}, \frac{x_2}{x_2}\right) = f(\bar{x}, 1) = f(\bar{x}) \tag{8.5}$$

in which $\bar{x}$ is the number of units x_1 per hectare and y/x_2 is the yield per hectare. Hence, the final production model is given by:

$$z = f(\bar{x}) \tag{8.6}$$

whereby z is the yield per hectare as a function of the number of units of x_1 added per hectare. This means that under the given assumptions (homogeneous production function), the optimal amount of "other input"(x_1) per hectare is independent of the number of hectares.

The profit is given by:

$$\pi = x_2(p_y f(\bar{x}) - w_1\bar{x} - w_2 1) \tag{8.7}$$

If the profit is maximised with regard to $\bar{x}$ by differentiating π and setting the derivative equal to zero, the condition for profit maximisation is given as:

$$p_y MPP_{\bar{x}} = VMP_{\bar{x}} = w_1 \tag{8.8}$$

which is in fact a point on the pseudo scale line for x_2 equal to 1

The equation for the pseudo scale line is found when:

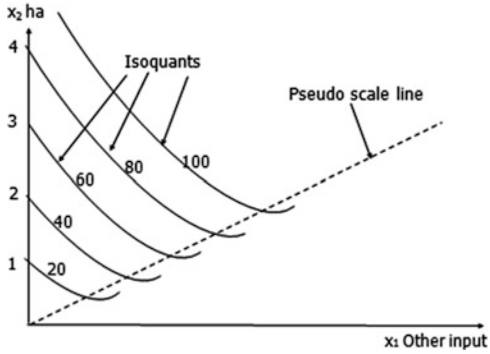


Fig. 8.2 Pseudo scale line

$$p_y MPP_1 = p_y \frac{\partial f(x_1, x_2)}{\partial x_1} = w_1 \tag{8.9}$$

Let us look at a specific example. As f is assumed to be homogeneous of degree one, it can be written as:

$$f(x_1, x_2) = g\left(\frac{x_1}{x_2}\right)x_2 \tag{8.10}$$

and if we further assume that g is a quadratic function given by:

$$g\left(\frac{x_1}{x_2}\right) = a + b\left(\frac{x_1}{x_2}\right) - c\left(\frac{x_1}{x_2}\right)^2 \tag{8.11}$$

in which a, b, and c are parameters, then the right hand side in Eq. 8.11 can now be inserted in Eq. 8.10, and if f is then differentiated with regard to x_1, MPP_1 is:

$$MPP_1 = \frac{\partial f}{\partial x_1} = b - 2cx_1x_2^{-1} \tag{8.12}$$

If this expression is inserted in Eq. 8.9, the equation for the pseudo scale line is given by:

$$\frac{x_1}{x_2} = \frac{p_y b - w_1}{p_y 2c} \tag{8.13}$$

which constitutes a straight line through the zero point, as illustrated in Fig. 8.2.

References

Balk, B. M. (1998). *Industrial price, quantity, and productivity indices*. Boston: Kluwer Academic Publishers.

Chambers, R. G. (1988). *Applied production analysis: A dual approach*. New York: Cambridge University Press.

9 The Company's Supply Function

9.1 Introduction

As described in Chap. 4, the company maximises its profit (profit maximisation) if production is expanded to the point where the marginal cost (i.e. the incremental cost of producing one more unit) is precisely equal to the product price. The product price is in fact equal to the additional revenue achieved from selling one more product unit. The criterion for profit maximisation can therefore also be expressed as the point where the marginal cost is equal to the marginal revenue.

The marginal revenue – i.e. the additional revenue achieved by selling one more unit – is not necessarily equal to the product price. Under special market conditions, the additional revenue achieved will be less than the price because the price decreases with an increase in sales. If this is the case, the units already being sold should be taken into consideration, as the price decrease in such a case will also affect the revenue from the sale of these units. Such situations, when the product price depends on the amount produced and sold, will be analysed in further detail in Chap. 13.

The present chapter is still based on the assumption that the company can produce and sell any (even large) amounts at the same price (perfect competition). Based on this, it is shown in the following that the company's supply of a product y can be derived from the cost function.

9.2 The Supply Curve

The criterion for profit maximisation when addressing the optimisation problem from the cost side has previously been derived in Chap. 5 (see Eq. 5.18). The optimal production is found when the output price equals the marginal costs.

In Fig. 9.1, the lower part of the previously derived Fig. 5.3 is repeated. The optimal production at different prices is illustrated by the points a, b, c, and d. As can be seen, the marginal cost curve (MC) in fact shows the relationship between

S. Rasmussen, *Production Economics*, Springer Texts in Business and Economics,
DOI 10.1007/978-3-642-30200-8_9,

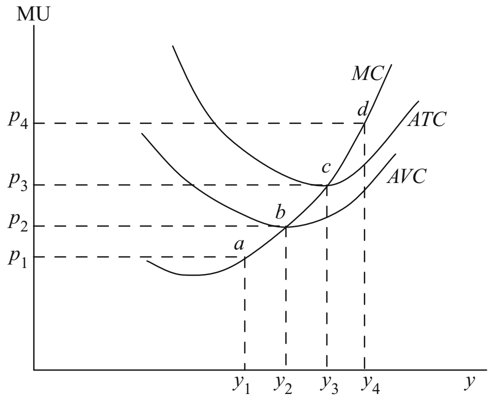


Fig. 9.1 The company's supply function

the price and the produced (and thereby the supplied) amount. However, the relationship between the price and the produced (supplied) amount is precisely the definition of a supply function. *Hence, the marginal cost curve is equal to the company's supply function or supply curve.*

The supply function is, however, only *part of the marginal cost curve*. Presume e.g. that the output price is p_1. A price of p_1 does not provide for cost coverage as point *a* is situated lower than the average cost. From an overall perspective, the company will, in such cases, produce at a loss and a rational company owner would, thus, not produce or supply anything at this low price p_1.

If the price is between p_2 and p_3, sales revenue per unit of output which is higher than the average variable costs (*AVC*) is achieved. The company owner will thus achieve a positive gross margin, i.e. a positive revenue to the coverage of (a part of) the fixed costs. And as the fixed costs per definition are fixed in the short run, production will be better than no production in the short run. However, in the long run, prices between p_2 and p_3 will not be sufficient for a profitable production. In the long run, there should also be coverage for the fixed costs (which are also variable in the long run), and at this price level the production will therefore gradually subside with the depreciation of the fixed assets.

If the price is higher than p_3, sales revenue per unit which is higher than the average total costs (*ATC*) is achieved. Hence, the company owner will achieve complete cost coverage and, in addition, an actual positive profit per unit, corresponding to the distance between the *MC* curve and the *ATC* curve. Hence, with this higher price it is particularly beneficial to produce, and production will take place – even in the long run – as the fixed costs (which will also be variable in the long run) are also covered.

Therefore, when defining the supply curve, it is important to differentiate between the short and the long run. *In the short run, the company's supply curve is the part of the marginal cost curve that is above the average variable cost curve (to the right of point b). In the long run, the company's supply curve is* *the part of the marginal cost curve that is above the average total cost curve (to the right of point c).*

As described here, it will never be profitable for a company that produces and sells under perfect competition to produce when the price is lower than the average variable costs (e.g. below p_2 in Fig. 9.1). By comparing with the derivation in Fig. 5.3 and the related production function on which it is based (Fig. 5.1), it can be seen that it will never be profitable to produce to the left of point *A* on the left hand side of Fig. 5.1. *As long as productivity is increasing, production should therefore be increased, and the optimal production is found at the part of the production function where productivity is diminishing*.

9.3 Adjustment in the Long Run

Each individual company's adjustment as described in Fig. 9.1 has some interesting macroeconomic implications.

An industry such as farming has traditionally been described as an industry under perfect competition.[1] Let us presume that all companies within the industry have an identical cost function, corresponding to the one shown in Fig. 9.1. Let us, furthermore, presume that the price initially is lower than p_3. In the long run, the industry's total supply will decrease with the wearing out of the companies' fixed assets under such conditions. As it is, there is no incentive for making new investments. The implication of the decreasing supply will be – everything else being equal – that the price will increase. When the price has increased to p_3 (or higher) there will no longer be any incentive to reduce production.

Let us instead presume that the price initially is higher than p_3. Each individual company achieves a profit, new companies are attracted and existing companies invest and expand production. The total effect is that the total supply of the industry increases, which – everything else being equal – results in a price decrease. When the price decreases to p_3 (or below) there will no longer be any incentive to expand production or to set up a company in the industry, as there is no longer any prospect of a positive profit.

Key Result

> The implication of the above is that, within any industry under perfect competition, there is a tendency for the price to move towards an equilibrium price corresponding to p_3 where complete cost coverage is achieved, and where productivity is the highest, and the returns to scale equal 1. Assumptions about companies in industries with perfect competition producing with constant returns to scale can, thus, be substantiated by the adaptation mechanism described here.

[1] There are segments of the industry in which this is no longer the case.

9.4 Derivation of the Supply Function. An Example

In Chap. 5 we derived the variable cost function (5.13) for a two input, Cobb-Douglas production function originally used in example 4.1 in Chap. 4. To show an example of the relationship between the production function, the cost function and the supply function, let us use the same production function as in example 4.1 and the following parameter values: $A = 6$; $b_1 = 0.3$; $b_2 = 0.5$. This means that the production function has the specific form, $y = 6x_1^{0.3}x_2^{0.5}$. Let us further assume that the input prices are $w_1 = 1$ and $w_2 = 2$. Then by inserting these parameter values in the variable cost function (5.13), we get the following variable cost function:

$$VC(y) = 0.31825y^{1.25}$$

By adding the fixed cost (FC) we get the total cost (TC). Taking the derivative of TC with respect to y we get marginal cost (MC). Dividing the variable cost by y we get average variable cost (AVC), and dividing total cost (TC) by y we get average total cost (ATC). The formulas for each of these terms are given below:

$$TC(y) = VC(y) + FC = 0.31825y^{1.25} + FC$$

$$MC(y) = \frac{\partial TC}{\partial y} = 0.397813y^{0.25}$$

$$AVC(y) = \frac{VC(y)}{y} = 0.31825y^{0.25}$$

$$ATC(y) = \frac{TC(y)}{y} = 0.31825y^{0.25} + \frac{FC}{y}$$

If we assume that the fixed cost is 2, then we get the following graphical illustration of the marginal and average cost curves (see Fig. 9.2) for this example.

The supply function (the MC curve) cuts the long cost curve (ATC) around MU 0.75, which means that the long run supply function is the MC curve above the value MU 0.75. Thus, in the long run the producer would not continue production unless the product price is more the MU 0.75. But what about the short run? In the special case illustrated here (Cobb-Douglas production function with the

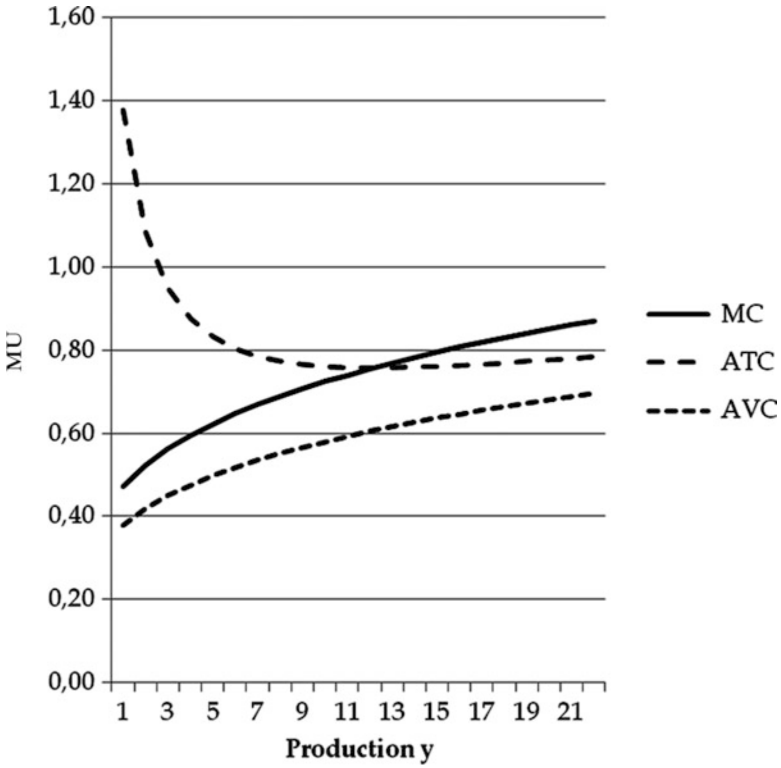


Fig. 9.2 Cost curves

parameters and input prices as stated above), it always pays to produce in the short run, because the average variable cost curve (AVC) is always below the MC curve. Therefore, the short run supply function is the whole MC curve.[2]

[2] As a good exercise, I recommend that the student analyses what would happen to the supply function if the input prices increase (from the present level of MU 1 and MU 2, respectively). Would the supply function move up or down, or would the slope change?