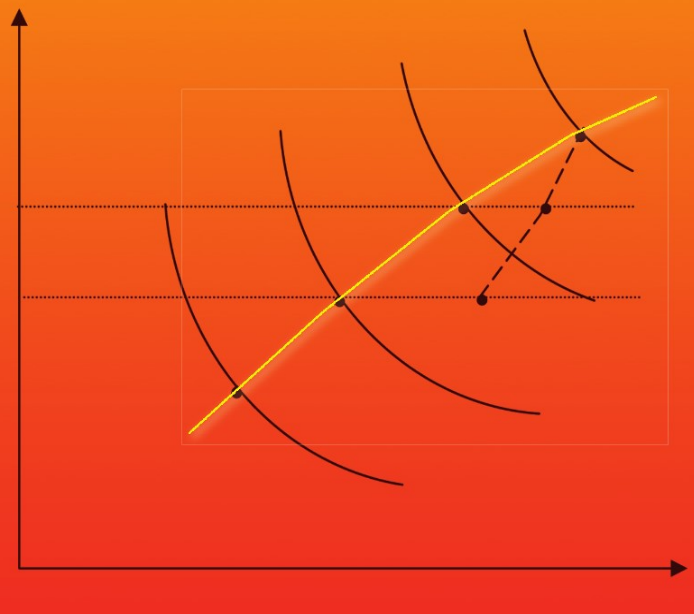
SVEND RASMUSSEN

Production Economics

The Basic Theory of Production Optimisation



Springer

Production Economics

Svend Rasmussen

Production Economics

The Basic Theory of Production Optimisation

Springer

Dr. Svend Rasmussen
Institute of Food and Resource Economics
University of Copenhagen
Rolighedsvej 25
1958 Frederiksberg C
Denmark
e-mail: sr@life.ku.dk

ISBN 978-3-642-14609-1 e-ISBN 978-3-642-14610-7

DOI 10.1007/978-3-642-14610-7

Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2010934762

Cover design: WMXDesign GmbH

Printed on acid-free paper

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

Preface

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@life.ku.dk

Copenhagen, April, 2010 Svend Rasmussen

Contents

Chapter 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.

First, the *factors of production* (also called input) 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 output).

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;
e. has the power to direct the activities of the suppliers of productive services, subject to the terms and conditions of their contracts;

S. Rasmussen, *Production Economics*, DOI: 10.1007/978-3-642-14610-7_1,

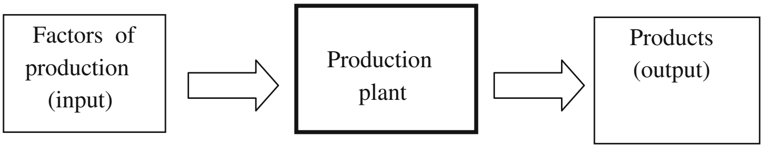


Fig. 1.1 Illustration of production

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 to 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 text-books, 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.

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 following 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 and 20 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 Chap. 20, is a powerful tool developed after the Second World War 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 operationalise 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. Pearson education* (3rd ed.). UK: Harlow.

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

Chapter 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 for 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

S. Rasmussen, *Production Economics*, DOI: 10.1007/978-3-642-14610-7_2,

Table 2.1 Yield with increased N addition

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

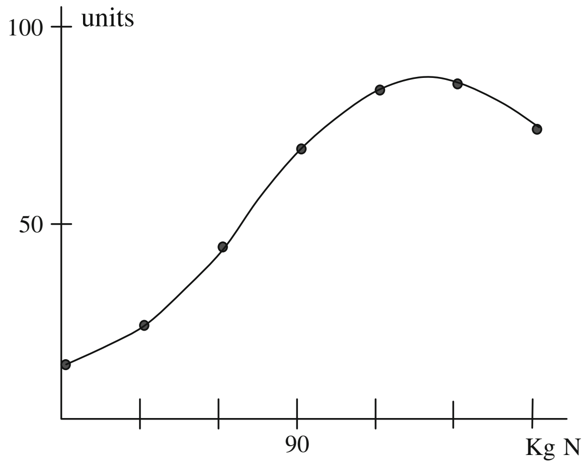


Fig. 2.1 Production function

choice of the functional form to illustrate the empirically observed relationships 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.

First, 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.

Second, 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 the amount of input and y 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:

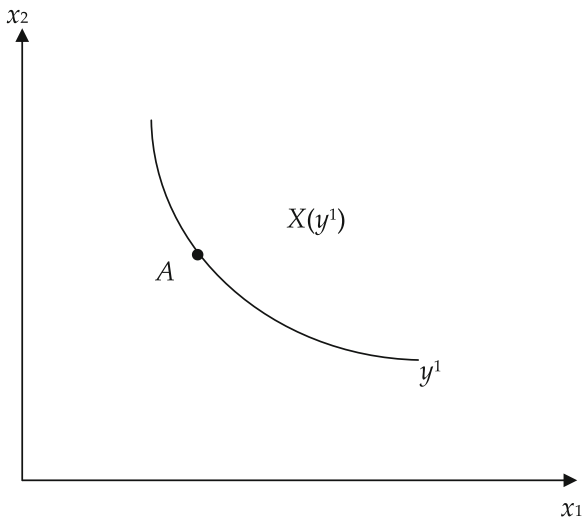
$$X(y) = \{x : x \, \text{can produce} \, y\} \tag{2.2}$$

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. $X(70) = \{x\text{: } x \geq 90\}$. 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.

Fig. 2.2 Isoquant and input set



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:

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

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. $Y(60) = \{y\colon y \leq 45\}$.

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.

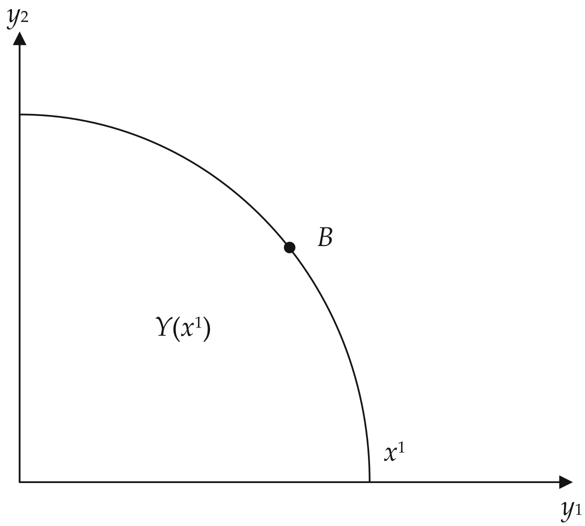


Fig. 2.3 Production possibility curve and the output set

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:

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

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:

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

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:

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

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. $\mathrm{d}f(x)/\mathrm{d}x$, 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.

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

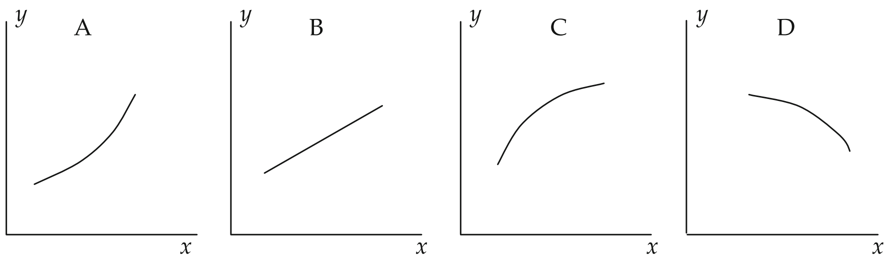


Fig. 2.4 Alternative production function shapes

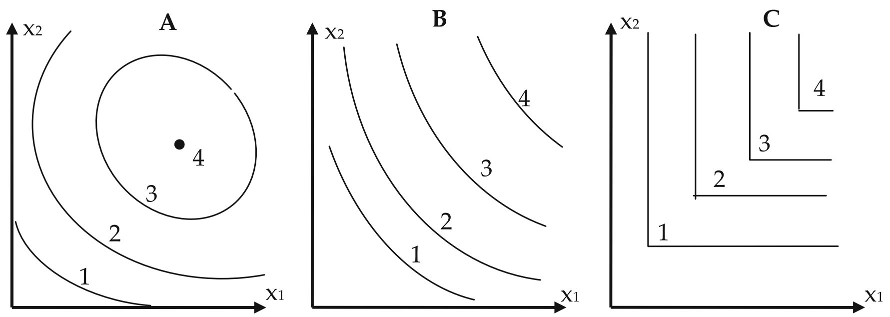


Fig. 2.5 Alternative sets of isoquants

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.

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, (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 $y = f(x_1|x_2,\ldots, x_n)$. 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 $y = f(x_1, x_2,|x_3,\ldots, x_n)$ 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, 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 (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).

Chapter 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:

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

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:

$$\text{APP} = f(x)/x = y/x.$$

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

[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*, DOI: 10.1007/978-3-642-14610-7_3,

3.2 Optimisation from the Input Side

When the optimal supply of input x is to be decided, 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

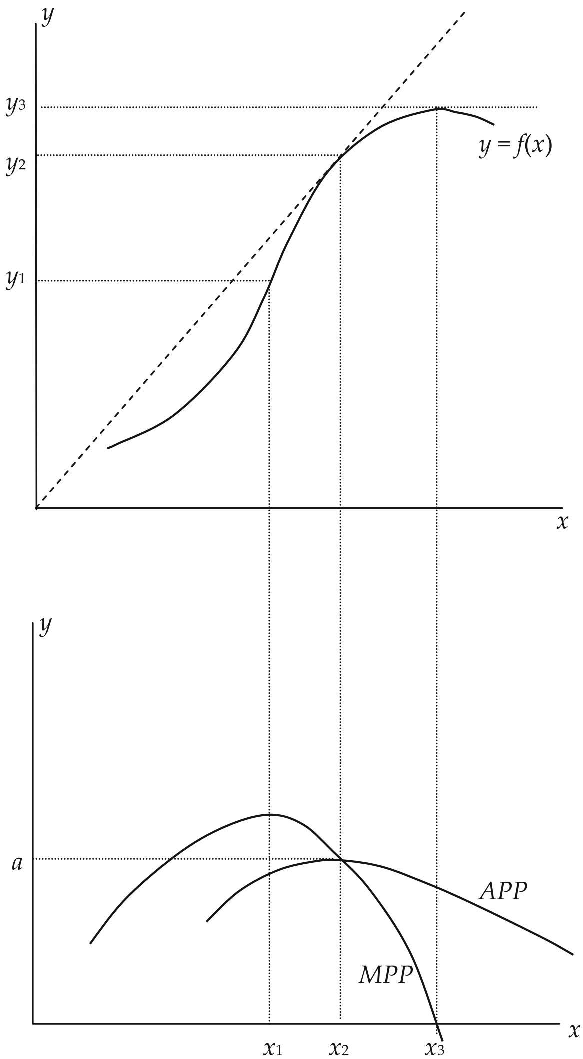


Fig. 3.1 The production function and MPP and APP

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 ($TR - TFC < 0$).

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 ($TR - TFC > 0$). 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:

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

in which p is the output price and w 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 (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:

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

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

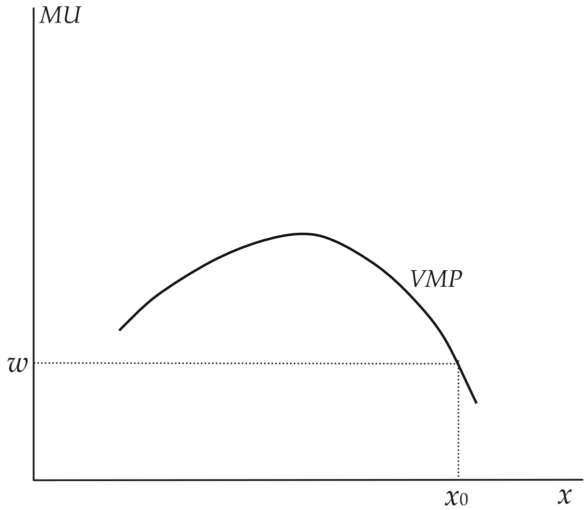


Fig. 3.2 Optimal input supply

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 $= \mathrm{d}f(x)/\mathrm{d}x = 0.8 - 0.01\mathrm{x}$
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 (3.2) the optimal condition can also be written as:

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

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 (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 (3.3) that if the input price increases compared to the output price, then the optimal input amount is

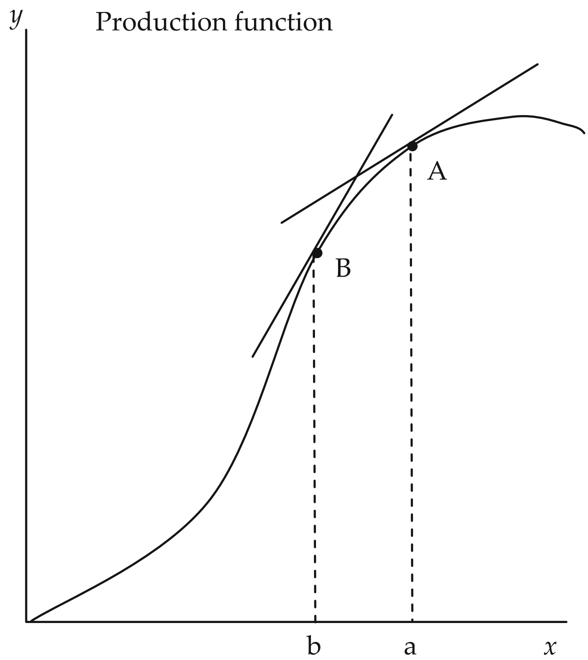


Fig. 3.3 Optimisation of input

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.

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 (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 (3.2), results in the sufficient condition for optimal input supply.

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

According to (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.

x	y (=$f(x)$)	MPP (exact) $\mathrm{d}f(x)/\mathrm{d}x$	MPP (approximated) $\Delta y/\Delta x$	VMP (approximated) 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 (3.1).

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

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 less than 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 (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 (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 (3.5) can now be expressed as:

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

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 (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:

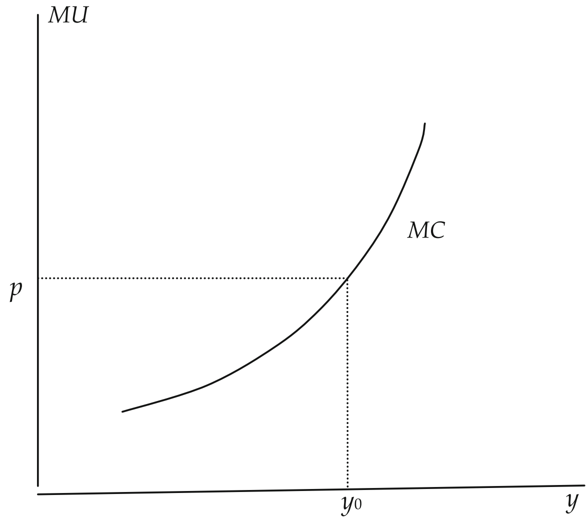
$$p = \text{d}c(w, y)/\text{d}y \equiv \text{MC} \tag{3.8}$$

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,

Fig. 3.4 Optimal production of y



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.

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.

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

y	Costs c $(w \times x^*(y))$	MC (approximated) $\Delta c/\Delta y$
23.5	50	
		7.69
30.0	100	
		9.09
35.5	150	
		11.11
40.0	200	
		14.29
43.5	250	
		20.00
46.0	300	
		33.33
47.5	350	
		100.00
48.0	400	

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 input application in this interval is $350 < x < 400$.

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: first, 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, 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.

Chapter 4
Production and Optimisation with Two or More Inputs

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*, DOI: 10.1007/978-3-642-14610-7_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:

$$\min_{x_1, x_2} \{w_1 x_1 + w_2 x_2\} \tag{4.1a}$$

under the constraint that:

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

The problem (4.1a, 4.1b) consists of minimising the variable costs (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:

$$L = w_1 x_1 + w_2 x_2 + \lambda(y - f(x_1, x_2)) \tag{4.2}$$

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 (4.1a, 4.1b):

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

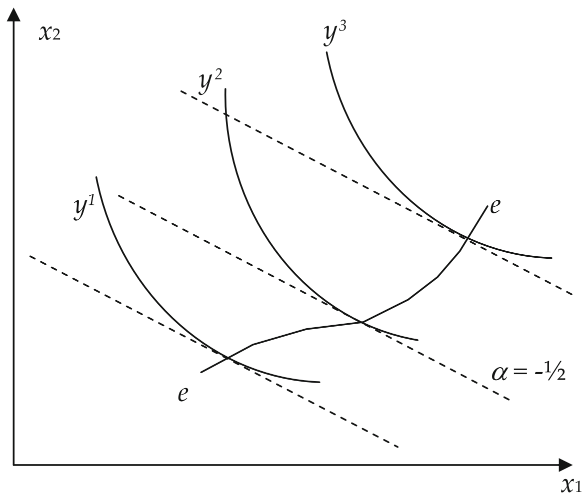


Fig. 4.1 Isoquants and expansion path

$$w_2 = \lambda \cdot MPP_2 \quad (4.3b)$$

$$y = f(x_1, x_2) \quad (4.3c)$$

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

$$\frac{w_1}{w_2} = \frac{MPP_1}{MPP_2} \quad (4.4)$$

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:

$$C = w_1 x_1 + w_2 x_2 \quad (4.5)$$

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

Finding the solution to (4.5) for x_2 produces:

$$x_2 = -\frac{w_1}{w_2} x_1 + \frac{C}{w_2} \quad (4.6)$$

which is a straight line in the x_1–x_2 plane with the slope $-(w_1/w_2)$ and intersection point with the x_2 axis corresponding to 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:

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

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 (4.7) is set equal to zero and solved with regard to dx_2/dx_1 the following result is generated:

$$\frac{dx_2}{dx_1} = -\frac{MPP_1}{MPP_2} = -MRS_{12} \quad (4.8)$$

which is in fact the ratio between changes in x_1 and x_2 when producing the constant product amount y. According to (4.8), the slope of the isoquant (dx_2/dx_1) can thus be expressed by the negative ratio between the marginal products for the two inputs MPP_1 and MPP_2. This ratio is called Marginal Rate of Technical Substitution, or simply Marginal Rate of Substitution (MRS). 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 (4.5) for the given production y can now be illustrated graphically, as shown in Fig. 4.2. According to (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 $-w_1/w_2 = -1/2$ 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) $8 = 1.8x_1^{-0.7}x_2^{0.5}$
(2) $12 = \lambda 3x_1^{0.3}x_2^{-0.5}$

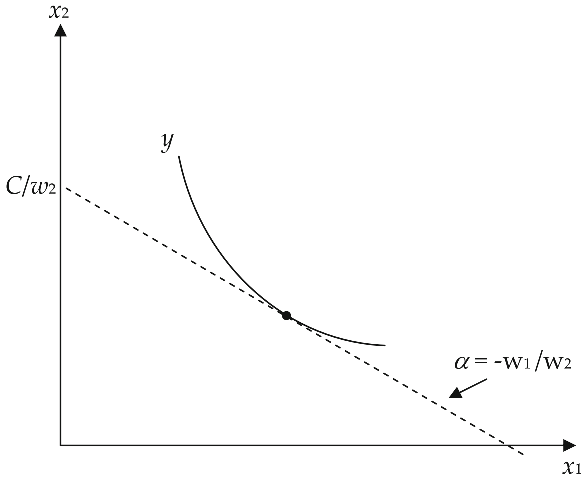


Fig. 4.2 Isoquant and isocost line

Dividing (1) by (2) will result in:

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

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: $x_2 = (24/21.6)x_1$.

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:

$$y = Ax_1^{b_1}x_2^{b_2}.$$

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

$$MPP_1 = Ab_1x_1^{b_1-1}x_2^{b_2}$$

and

$$MPP_2 = Ab_2x_1^{b_1}x_2^{b_2-1}$$

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

$$\frac{w_1}{w_2} = \frac{b_1x_2}{b_2x_1},$$

which can also be expressed as:

$$x_2 = \frac{w_1b_2}{w_2b_1}x_1,$$

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 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:

$$y = t^{(b_1+b_2)}Ax_1^{b_1}x_2^{b_2} = t^n f(x_1, x_2).$$

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:

$$y = f(x_1, x_2) = a_0 + a_1x_1 + a_2x_2 - a_{11}x_1^2 - a_{22}x_2^2 + a_{12}x_1x_2$$

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 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 one 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 $\otimes$.

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 $\otimes$, 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 $\otimes$ 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.

4.4 Maximisation of Production Under Budget Constraint

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

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

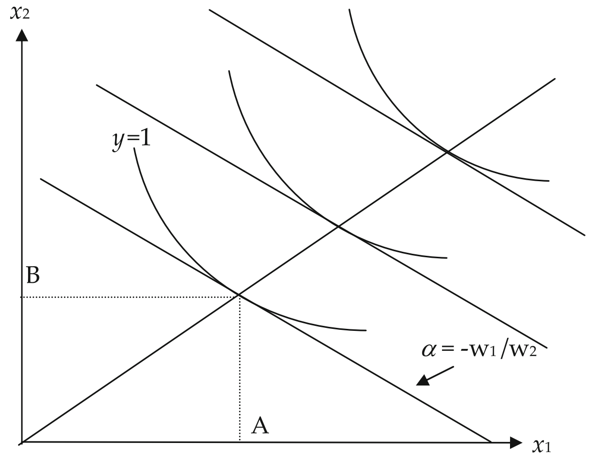
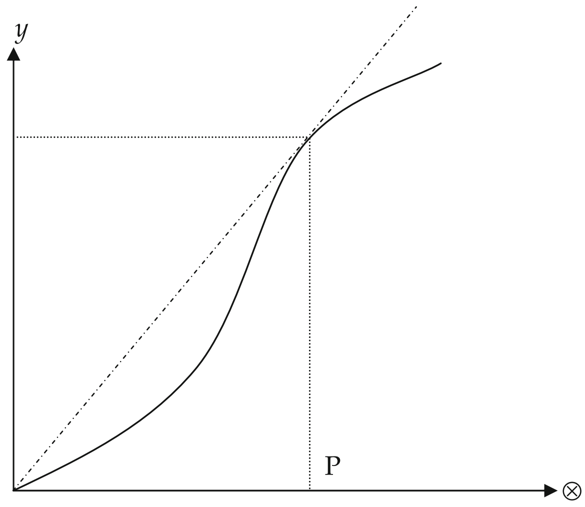


Fig. 4.3 Homothetic production function

Fig. 4.4 Homothetic production function



under the constraint that:

$$C = w_1x_1 + w_2x_2 \tag{4.9b}$$

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

$$L = f(x_1, x_2) + \theta(\mathrm{C} - (w_1x_1 + w_2x_2)) \tag{4.10}$$

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 (4.11a) by (4.11b) produces the necessary condition for the maximisation of (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 company if you wish 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 the profit maximum) 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:

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

Furthermore, (as it will soon turn out) precisely at the profit-maximising point A, the two fractions in (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 (4.16) can be derived by maximising the profit as the function of the two inputs:

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

[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.

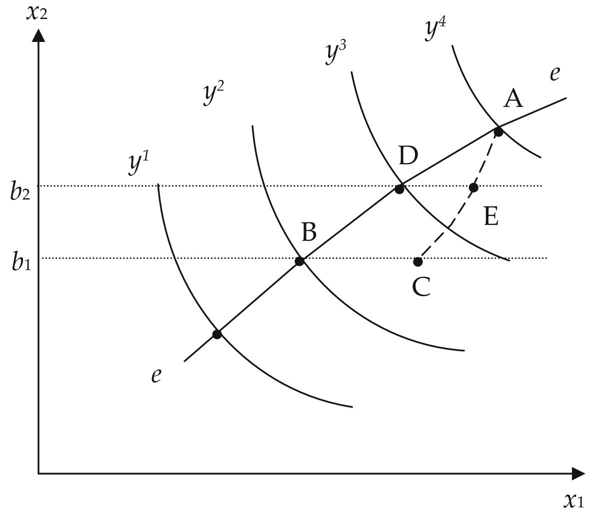


Fig. 4.5 Isoquants and profit maximisation

Differentiating the profit in (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*:

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

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

which in fact corresponds to the criteria in (4.15) and (4.16). Hence, the previously derived result regarding one input is (naturally) also true for two inputs, i.e. that each input must be added in an amount 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:

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

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 (4.4). Please note that (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, (4.4a) can also be written as:

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

Comparing (4.4b) with (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 (4.4a) is higher than for the corresponding fractions in (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 (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. 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 (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 (4.22).

If the same analysis is carried out for all possible fixed levels of the input x_2, Fig. 4.5 will show 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 with 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 (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

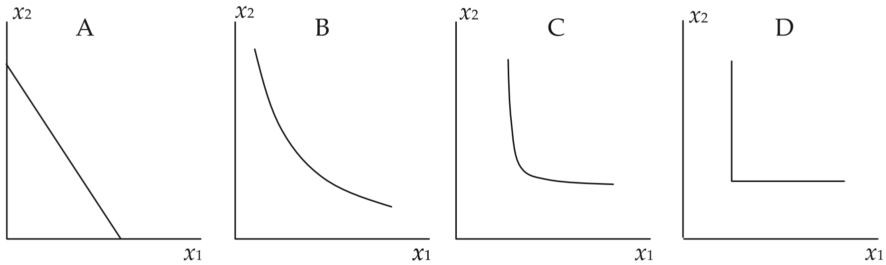


Fig. 4.6 Alternative shapes of isoquants

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 percentage of change in one expression through the percentage of change in another expression. *The input elasticity of substitution* (ε_{sh}) originally proposed by Heady (1953) can be approximated for small changes (Δx) by:

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

but is more formally *defined* as:

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

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.

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. 27 ff.).

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.

Chapter 5
Costs

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 will 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*, DOI: 10.1007/978-3-642-14610-7_5,

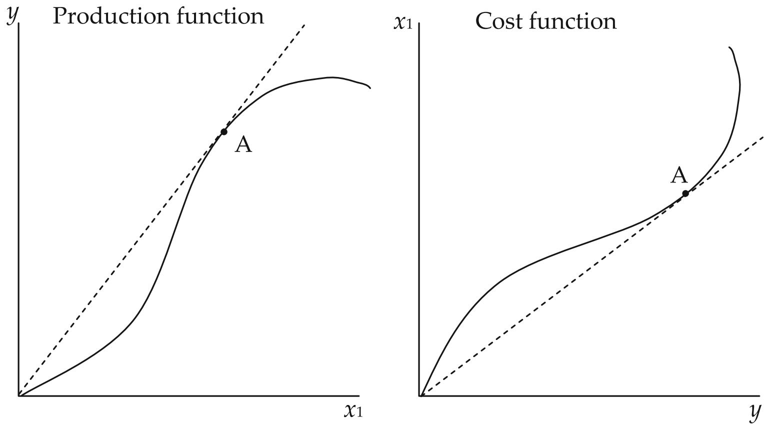


Fig. 5.1 Cost function as a mirror reflection (dual) of the production function

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 $y = f(x_1|x_2, \ldots, x_n)$, whereby the inputs $x_2, \ldots, x_n$ 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 costsTC* 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:

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

Average total costs:

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

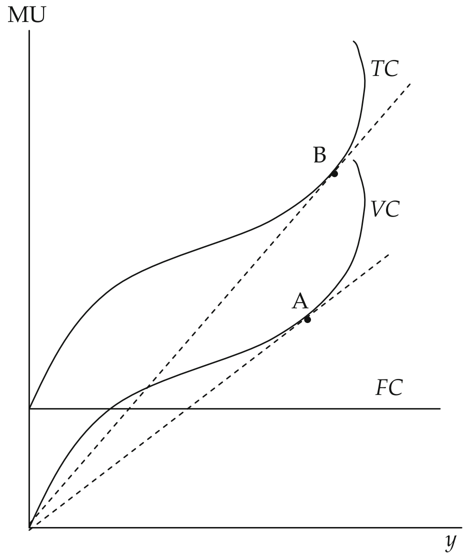


Fig. 5.2 Fixed, variable, and total costs

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.

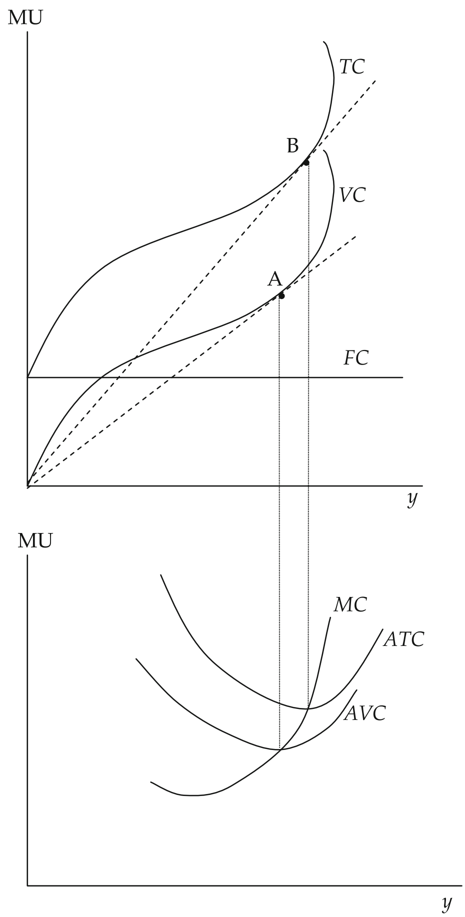


Fig. 5.3 Average costs and marginal costs

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 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*:

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

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

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

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:

$$x_2 = \left(\frac{y}{A}\right)^{1/(b_1+b_2)}\left(\frac{b_2w_1}{b_1w_2}\right)^{b_1/(b_1+b_2)} \tag{5.12}$$

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:

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

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 (5.13) has the form (5.14), where:

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

and $h(w_1, w_2)$ is the remainder of (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.

x_1	x_2	Costs C	Production y	Approximated marginal costs $\Delta C/\Delta y$
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

(continued)

(continued)

x_1	x_2	Costs C	Production y	Approximated marginal costs $\Delta C/\Delta y$
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 (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:

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

The formulation in (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, a similar condition is derived with the product price being dependent on the production y.)

Differentiating the profit in (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 (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 (5.18) generates the same result as with profit maximisation from the input side [see (4.19)]. Please note in this connection that the use of the cost function in (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.

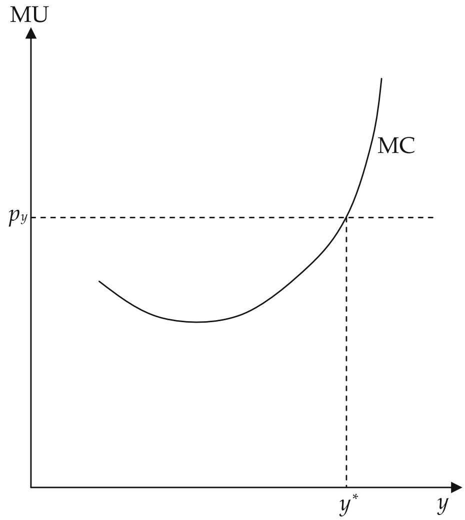


Fig. 5.4 Determination of optimal production

The optimisation criterion in (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:

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

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).

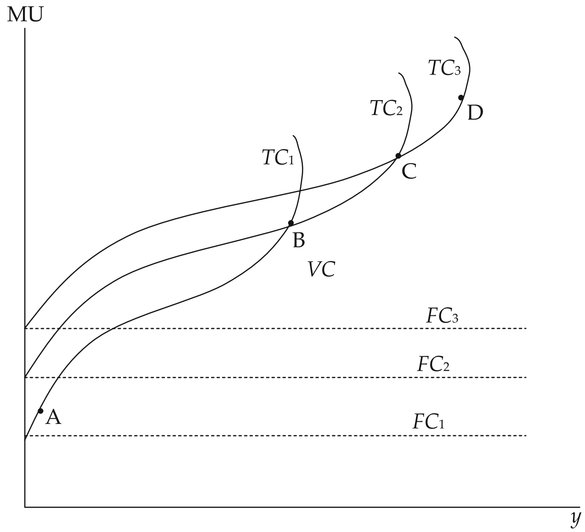


Fig. 5.5 Short run costs at different plant sizes

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.

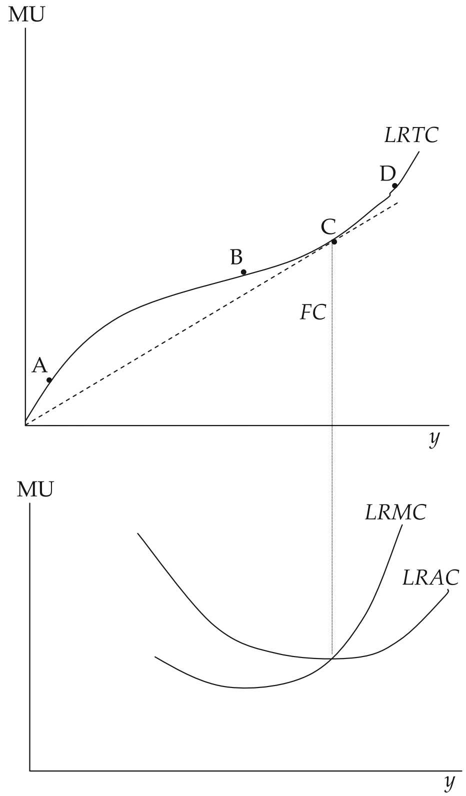
These conditions are outlined in Fig. 5.5, in which the curve TC_1 and FC_1 correspond to the original cost curves in the upper part of Fig. 5.3.

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.

Fig. 5.6 Long run costs



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 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* (vs. 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 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.

5.4.1 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).

5.4.2 The Accounting Principle

Calculation of the costs according to the accounting principle is based on *reacquisition* 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 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

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

accounting principle and its alternative versions are further discussed in the appendix *Profit concepts*.

You will find some examples below to illustrate these two principles:

Production factor	Costs	
	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 these 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 × 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 is 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.

Chapter 6
Productivity, Efficiency and Technological Changes

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*, DOI: 10.1007/978-3-642-14610-7_6,

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

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 *thechange in productivity* from period t to period $t + 1$ equals:

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

The last parenthesis in (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 to 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 below, 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:

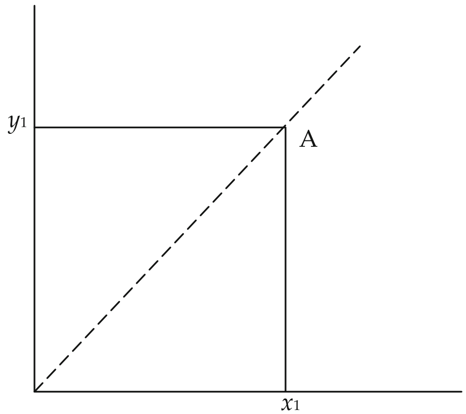


Fig. 6.1 Illustration of productivity

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

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:

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

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:

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

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:

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

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*.

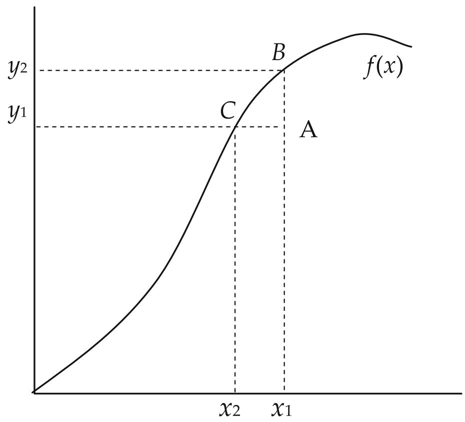


Fig. 6.2 Illustration of efficiency

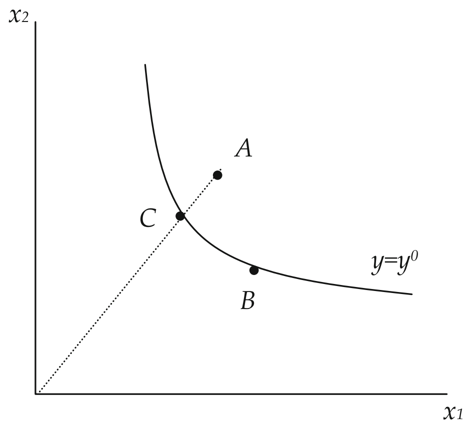


Fig. 6.3 Illustration of efficiency

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 for e.g. possible to produce the same amount in point C. The distance AC, or the distance $0C$ 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).

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 $(\tau - 1)$—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 $(\tau - 1)$, whereτ 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.

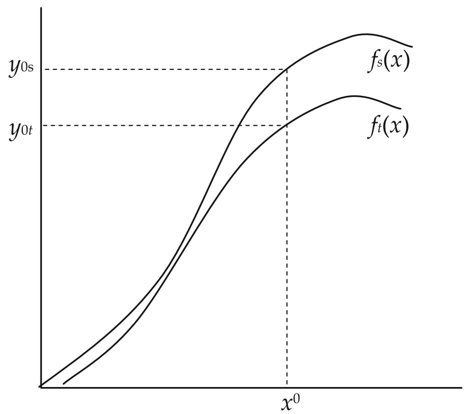


Fig. 6.4 Illustration of technological changes

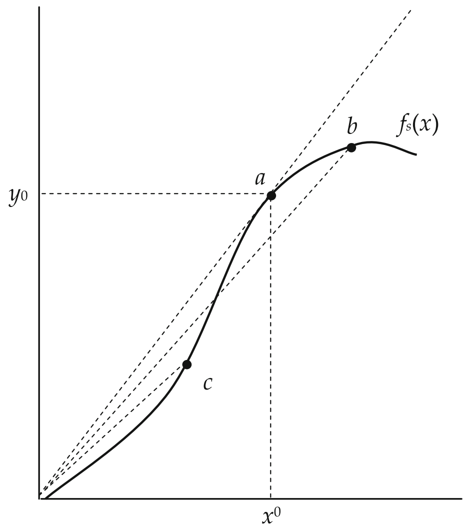


Fig. 6.5 Description of scale

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:

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

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.*

[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.

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 $0B$ than line $0A$. 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

Fig. 6.6 Productivity increase

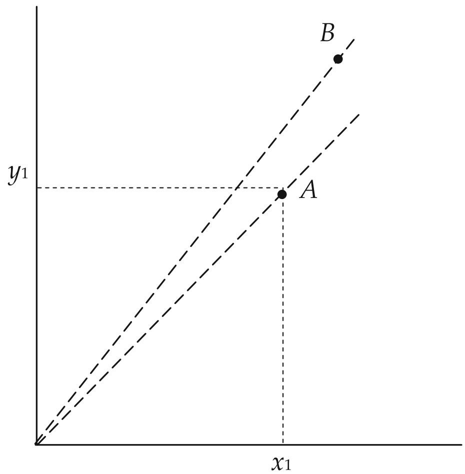


Fig. 6.7 Efficiency increase

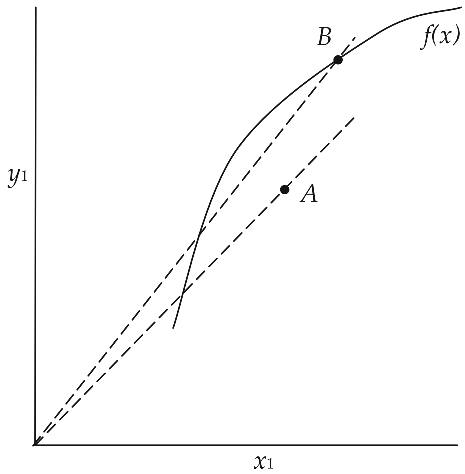
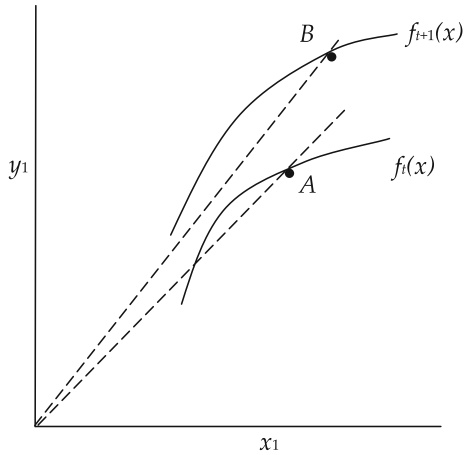


Fig. 6.8 Technological change



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).

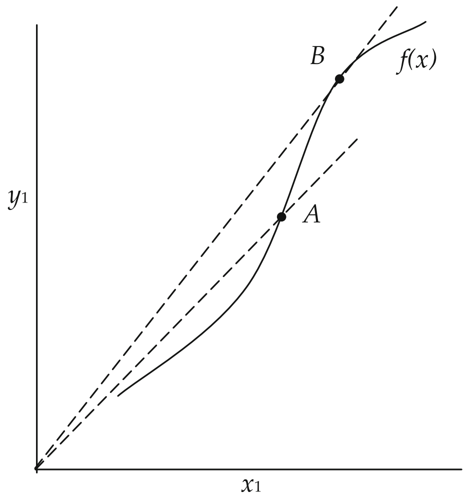
The productivity in period t, when the productivity in period s is set equal to 1, is:

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

The actual measured output y_t can be expressed as:

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

Fig. 6.9 Change of scale



in which τ_t is the expression for the efficiency in period t. The same is true for period s. Inserting (6.8) in (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 (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 whereby all inputs are changed by the same factor), the relation between input in the two periods can be written as $x_t = \kappa x_s$, whereby 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 (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 (6.11)], and technological changes [the last component in the right hand side of (6.11)], there is one additional component $\kappa^{(\varepsilon-1)}$ expressing the scale effect, as illustrated in (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.
Coelli, T., Prasada Rao, D. S., O'Donnell, C. J., & Battese, George. (2005). *An introduction to efficiency and productivity analysis* (2nd ed.). New York: Springer.

Chapter 7
Input Demand Functions

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, ..., 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, it could be imagined 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*, DOI: 10.1007/978-3-642-14610-7_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 (3.2)]:

$$\frac{p_y \mathrm{MPP}_1}{w_1} = \frac{\mathrm{VMP}_1}{w_1} = 1 \tag{7.2}$$

whereby 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 $\mathrm{VMP}_1 = w_1$ at the decreasing part of the VMP-curve (see Sects. 3.1, 3.2 in Chap. 3).

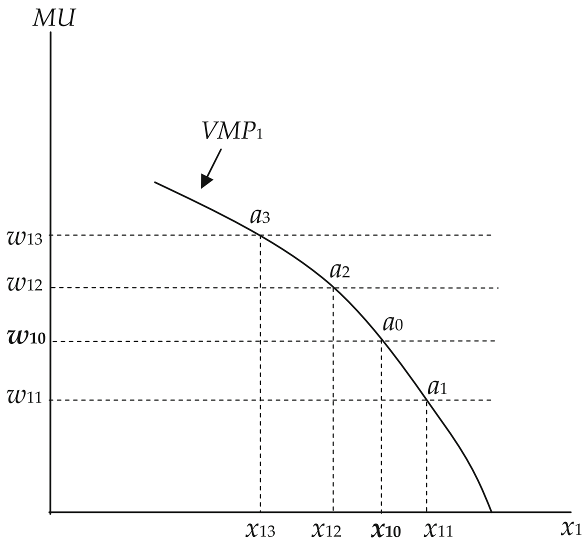
The criterion is illustrated graphically in Fig. 7.1. in which the initial price w_1 is presumed to be equal to w_{10}, and the optimal application of input x_1 therefore equals x_{10}.

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 $\mathrm{VMP}_1 = w_1$ can be written $p_y \mathrm{MPP}_1(x_1, \alpha) = w_1$. Solving for x_1 using the implicit function theorem provides the solution:

Fig. 7.1 Demand function for input



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

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 (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:

$$Y = f(x_1) = Ax_1^b. \tag{7.4}$$

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

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

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 (7.7), the input demand function can be expressed as:

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

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 falls with an increasing input price and rises with an increasing 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:

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

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 If you look at the previous example [see formula (7.7)] and calculate the own-price elasticity using the middle term in (7.9), (7.7) is differentiated first 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 (7.9). Taking the logarithm of x_1 in (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 (7.9) as:

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

Differentiating (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.

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

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.

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 (4.18a) and (4.18b)] for two variable inputs:

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

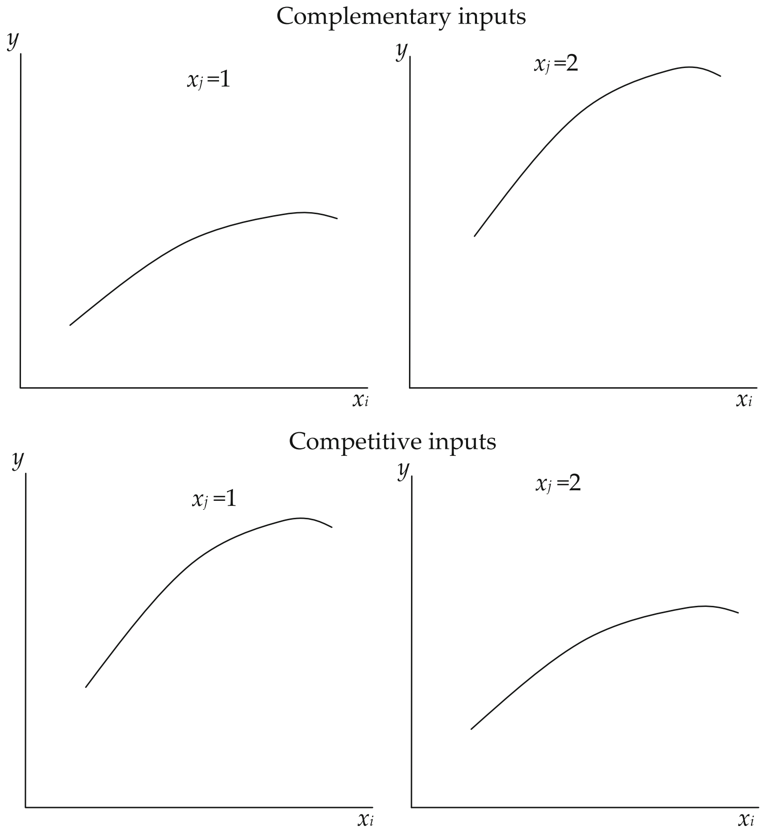


Fig. 7.2 Interaction between input *i* and input *j*

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

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 with 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).

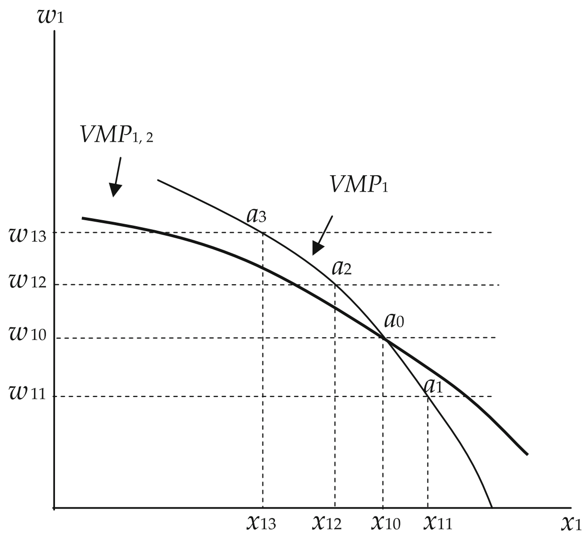


Fig. 7.3 Demand function for input x_1 with two variable inputs

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 (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 (7.16a) is:

$$ap_yAx_1^{a-1}x_2^b = w_1 \tag{7.19a}$$

and (7.16b):

$$bp_yAx_1^a x_2^{b-1} = w_2 \tag{7.19b}$$

Isolating x_1 in (7.19a) generates:

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

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 (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 (7.20) with regard to w_1 before the function-related relationship between x_2 and w_1 is established.The method for this is a simultaneous solution of (7.19a) and (7.19b). Dividing (7.19a) by (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 (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:

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

which is less than zero when $a + b$ is less than 1, i.e. when the returns to scale is decreasing (see Chap. 4). Hence, the own-price elasticity for input, when the production function is a Cobb–Douglas function, is negative when the returns to scale is decreasing. Comparing (7.23) with (7.13) shows that the demand elasticity for input x_1 depends on whether there are other variable inputs. As mentioned in the above, 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

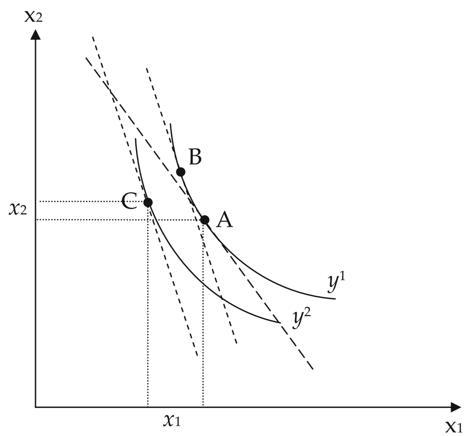


Fig. 7.4 Substitution effect and income effect

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:

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

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 of 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

[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.

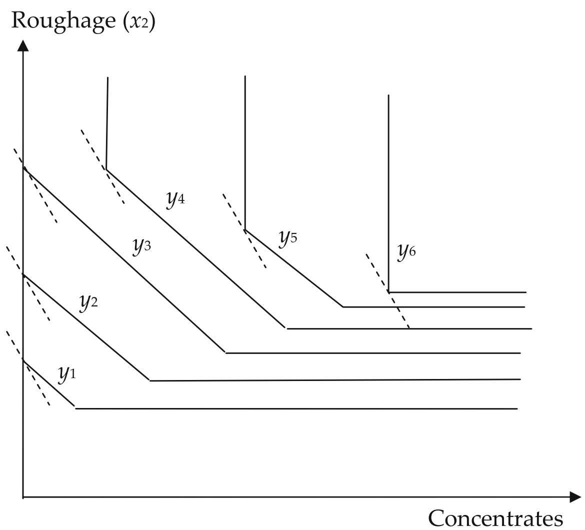
production from y^2 (point C) to y^1 (point A) entails that the consumption of x_2 in fact decreases (however, the consumption of x_1 increases in return).

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 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.

Fig. 7.5 Isoquants for milk production



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 below. 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.

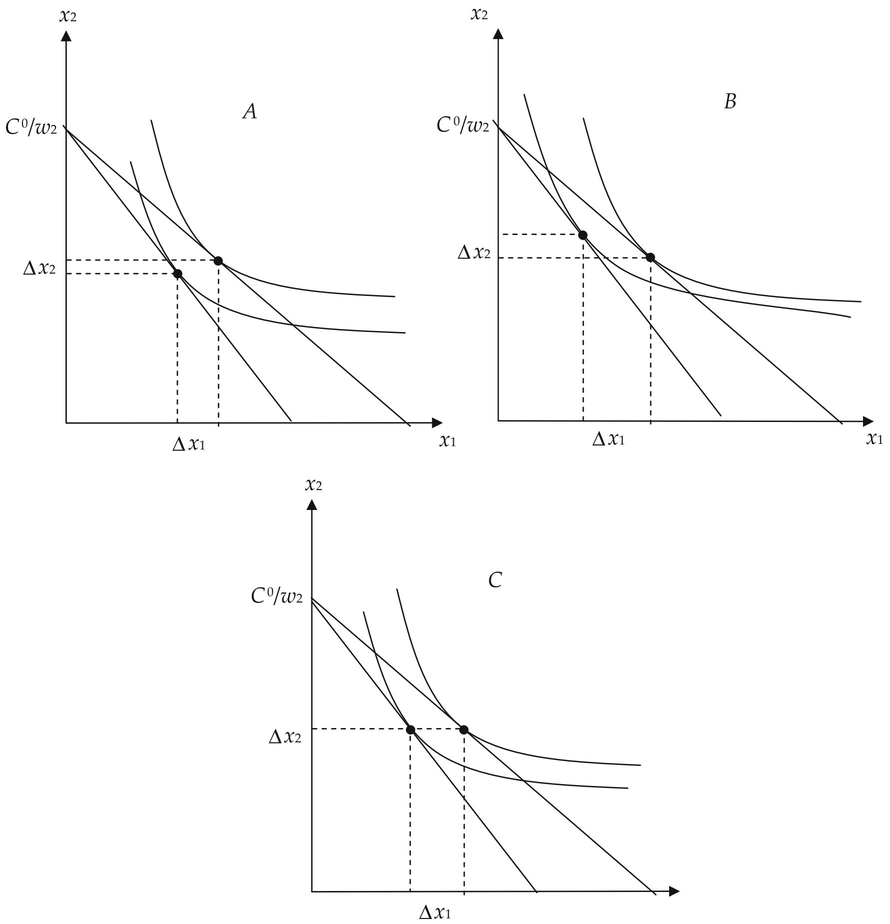


Fig. 7.6 Substitution under budget constraint

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.

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 (4.10), (4.11), and (4.12) in Chap. 4]:

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

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:

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

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

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

Dividing (7.26a) by (7.26b) produces the necessary condition for the maximisation of (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 (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 (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 (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.

Chapter 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 (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*, DOI: 10.1007/978-3-642-14610-7_8,

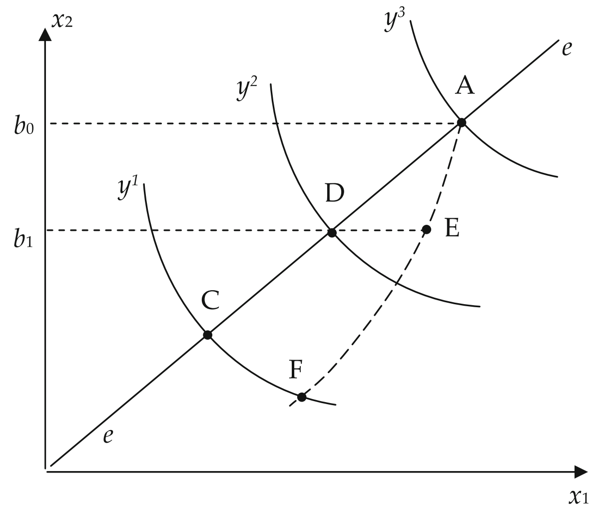


Fig. 8.1 Isoquants and pseudo-scale line

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

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

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}, … , x_{1p}, "produces" the (intermediate) "product" x_1 which is then used as an input in the final production function f in (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 (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:

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

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 for 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 MRS (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), Chapter 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 f in (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 (8.4), which means that (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:

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

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:

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

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:

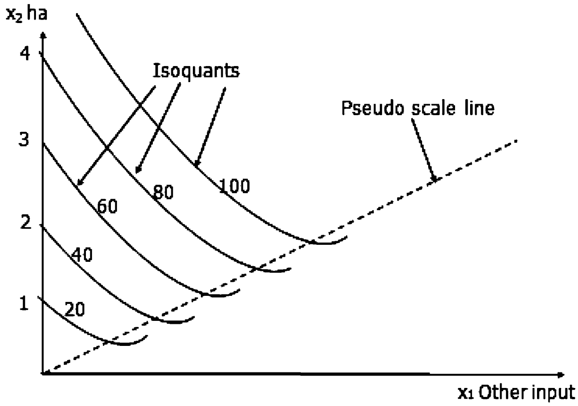


Fig. 8.2 Pseudo scale line

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

in which a, b, and c are parameters, then the right hand side in (8.11) can now be inserted in (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 (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.

Chapter 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 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

S. Rasmussen, *Production Economics*, DOI: 10.1007/978-3-642-14610-7_9,

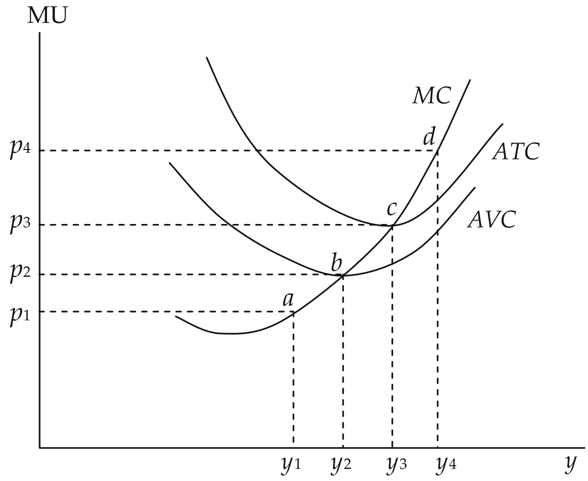


Fig. 9.1 The company's supply function

can be seen, the marginal cost curve (MC) in fact shows the relationship between 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 (AVCs) is achieved. The company owner will thus achieve a positive gross margin, i.e. a positive input 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 (ATCs) 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 AVC 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 ATC 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 AVCs (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.

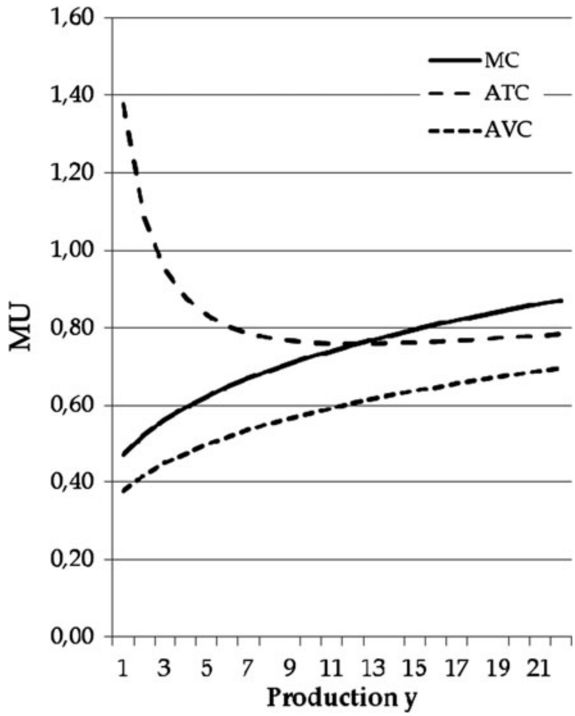
9.4 Derivation of the Supply Function: an Example

In Chap. 5 we derived the variable cost function (Eq. 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:

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

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 AVC and dividing total cost (TC) by y we get ATC. The formulas for each of these terms are given below:

Fig. 9.2 Cost curves



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

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

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

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

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 parameters and input prices as stated above), it always pays to produce in the short run, because the 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?

Chapter 10
Optimisation of Production Under Restrictions

10.1 Introduction

The general criteria for the adjustment of production when product and factor prices change were described in Chap. 3–5. The general results are derived under the general assumption that the producers wish to maximise profit. The general criterion for profit maximisation is that the addition of input should be continued as long as the increase in total revenue (i.e. marginal revenue) is greater than the increase in costs (i.e. marginal costs). This general criterion includes the following two criteria as special cases: (1) The value of the marginal product for all (variable) inputs must be equal to the price of the corresponding input ($VMP_i = w_i$), and (2) The marginal costs must be equal to the price of output ($MC = p_y$).

In practice, it is often not possible to adjust the production freely. There will be cases when, for some reason or other, it is not possible to buy and use the desired amount of input, or when it is not possible to produce the desired amount of output. Examples of this were already outlined in Chap. 4 with the discussion of the optimisation of the production with two (or more) inputs under budget constraints, i.e. under conditions where there is not enough money to buy all the input desired to maximise profit. In this connection, we saw that the criterion for profit maximisation included another criterion, namely the minimisation of costs. In connection with the minimisation of costs, it was concluded that the optimal usage of multiple inputs is achieved if two arbitrary inputs are combined so that:

$$\frac{w_i}{w_j} = \frac{MPP_i}{MPP_j} \text{ for any input } i \text{ and } j.$$

More generally, this chapter will discuss the problems involved in the adjustment of the production under restrictions and will contain examples of the analysis of such problems.

S. Rasmussen, *Production Economics*, DOI: 10.1007/978-3-642-14610-7_10,

10.2 General Method for Optimisation Under Restrictions

The problem of maximising the profit under restrictions can be formally formulated as the following optimisation problem:

$$\underset{x_1,\ldots,x_n}{\text{Max}}\ \{\pi(x_1,\ldots,x_n)\} = \text{Max}\ \{p_y f(x_1,\ldots,x_n) - w_1x_1 - \cdots - w_n x_n\} \tag{10.1a}$$

under the constraint:

$$g(x_1,\ldots,x_n) = 0 \tag{10.1b}$$

where the function g expresses a restriction, e.g. a budget constraint, as shown above in Chap. 4. A budget constraint has the form: $C = w_1x_1 + \cdots + w_n x_n$, and, written as in the form (10.1b), can be expressed as: $g(x_1,\ldots,x_n) = C - w_1x_1 - \cdots - w_n x_n = 0$.

Maximisation of a function under constraints of a form such as in (10.1b) can be done by determining the set of x's and the shadow price λ which maximises the Lagrange function L given by the following function[1]:

$$L = \pi(x_1,\ldots,x_n) + \lambda g(x_1,\ldots,x_n) \tag{10.2}$$

In general, a Lagrange function L is created by taking the object function (the function that is to be optimised (maximised or minimised)—in this case, the profit), and then adding the constraint g, multiplied by a Lagrange multiplier—called λ here.

The maximisation of L is carried out in the usual way by taking the partial derivatives and setting them equal to zero. This produces the following $n + 1$ condition for optimum:

$$\frac{\partial \pi}{\partial x_1} + \lambda \frac{\partial g}{\partial x_1} = 0 \tag{10.3.1}$$

$$\vdots$$

$$\frac{\partial \pi}{\partial x_n} + \lambda \frac{\partial g}{\partial x_n} = 0 \tag{10.3.n}$$

$$g(x_1,\ldots,x_n) = 0 \tag{10.3.n + 1}$$

With the budget constraint $C - w_1x_1 - \cdots - w_n x_n = 0$ as the example of the constraint, the Lagrange function L can be written as:

$$L = \pi(x_1,\ldots,x_n) + \lambda(C - w_1x_1 - \cdots - w_n x_n) \tag{10.4}$$

[1] The Lagrange function is a well known mathematical tool used in economics. For further explanation see any advanced mathematical textbook or mathematical economics textbook such as Chiang (1984).

Presupposing, for the sake of simplification, that $n = 2$ (only two variable inputs), then the optimisation conditions (10.3.1–10.3.n+1) can be expressed as:

$$p_y MPP_1 - w_1 - \lambda w_1 = 0 \tag{10.5.1}$$

$$p_y MPP_2 - w_2 - \lambda w_2 = 0 \tag{10.5.2}$$

$$C - w_1 x_1 - w_2 x_2 = 0 \tag{10.5.3}$$

which can furthermore be expressed as:

$$\lambda = \frac{VMP_1 - w_1}{w_1} \tag{10.6a}$$

$$\lambda = \frac{VMP_2 - w_2}{w_2} \tag{10.6b}$$

$$C = w_1 x_1 - w_2 x_2 \tag{10.6c}$$

As can be seen, this is a system consisting of three equations with the three unknowns x_1, x_2, and λ, and which therefore can (normally) be solved. The simplest way to find a solution to x_1 and x_2 is by dividing (10.5.1) by (10.5.2), thereby eliminating λ, and the remaining two equations are easily solved with regard to x_1 and x_2.

But what is the expression λ? In connection with the presentation of the Lagrange function (10.2) it was referred to as the *shadow price*. The reason for this is explained in the following.

Firstly, it is established that (10.6a) and (10.6b) contain two expressions of λ. If these two expressions are set equal to each other, the following familiar condition from Chap. 4 (see (4.4)) is generated:

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

expressing that the optimum can be found on the expansion path.

It is then established that if the producer is in fact capable of maximising the profit (i.e. that the restriction presented is *not* effective), then $VMP_1 = w_1$, and $(VMP_1 - w_1)/w_1$ in (10.6a) therefore equals 0 (zero) which means that $\lambda = 0$.

If the restriction, on the other hand, is effective (profit maximisation is *not* possible), then $VMP_1 - w_1$ is *greater* than 0, cf. the discussion in Sect. 4.5 (see (4.20)). If this is the case, λ is also positive and—as it will turn out—is in fact the expression of how much the profit could have been increased if there had been one more monetary unit (MU) available compared to the present budget constraint.

As can be seen, the numerator in (10.6a) is equal to the extra profit achieved by using one more unit of x_1. If this is divided by the price of x_1, i.e. by w_1, the result is an expression of the extra profit per MU added. However, this is in fact precisely what is expressed in (10.6a)—and thereby by λ! Hence, λ is an expression of the extra profit that would have been achieved with one more MU at disposal—i.e. if the budget had allowed the consumption of one more MU.

The following is generally true for Lagrange multipliers: The Lagrange multiplier is an expression of how much the value of the criterion function (or the object function—the function that should be maximised or minimised) will increase if the restriction in question was relaxed by precisely one unit. In the example here, where the aim is to maximise the profit, and the restriction comprises the amount of MUs at disposal for purchasing variable inputs, the Lagrange multiplier in fact expresses the extra profit, if there had been one more MU at disposal.

The Lagrange method can also be used for optimisation if there are *multiple restrictions*. In its general form, the Lagrange method can therefore be formulated as follows:

$$\underset{x_1,\ldots,x_n}{\text{Max}} \{\pi(x_1,\ldots,x_n)\} = \text{Max} \{p_y f(x_1,\ldots,x_n) - w_1 x_1 - \cdots - w_\text{n} x_\text{n}\} \tag{10.7}$$

under the constraints:

$$g_1(x_1,\ldots, x_n) = 0 \tag{10.8.1}$$

$$\vdots$$

$$g_p(x_1,\ldots, x_n) = 0 \tag{10.8.p}$$

in which the functions $g_1 \ldots g_p$ express the total of p restrictions, and where the Lagrange function L now has the following general form:

$$L = \pi(x_1,\ldots, x_n) + \lambda_1 g_1(x_1,\ldots, x_n) + \cdots + \lambda_p g_\text{p}(x_1,\ldots, x_n) \tag{10.9}$$

with the total of p Lagrange multipliers $\lambda_1 \ldots \lambda_p$. The solution is found by maximising L with regard to the total of now $n + p$ variables, as (10.9) is differentiated and the derivatives are set equal to zero.

Example 10.1 The point of reference is a previous example, namely Example 4.1 in Chap. 4 (see also Example 5.2 in Chap. 5). A production function of $y = f(x_1,x_2) = 6x_1^{0,3}x_2^{0,5}$ is presupposed. The price of input x_1 is 8 ($w_1 = 8$), the price of input x_2 is 12 ($w_2 = 12$), and the output price (p_y) is 7. The producer wishes to maximise profit. There are two restrictions: (1) a budget constraint C which entails that input can be bought at max MU 200, and (2) a restriction concerning the consumption of x_2, as the government has imposed the restriction that companies must consume precisely ten units of x_2.

The object function is:

$$\max\left[7 \times \left(6x_1^{0,3}x_2^{0,5}\right) - (8x_1 + 12x_2)\right]$$

The restrictions are:

$$\text{g}_1(x_1, x_2) = 200 - 8x_1 - 12x_2 = 0$$

$$\text{g}_2(x_1, x_2) = 10 - x_2 = 0$$

The Lagrange function is:

$$L = \left[7 \times \left(6x_1^{0,3}x_2^{0,5}\right) - (8x_1 + 12x_2)\right] + \lambda_1(200 - 8x_1 - 12x_2) + \lambda_2(10 - x_2)$$

Differentiating L produces the following four partial derivatives:

(1) $\partial L/\partial x_1 = 12.6x_1^{-0,7}x_2^{0,5} - 8 - 8\lambda_1$
(2) $\partial L/\partial x_2 = 21x_1^{0,3}x_2^{-0,5} - 12 - 12\lambda_1 - \lambda_2$
(3) $\partial L/\partial \lambda_1 = 200 - 8x_1 - 12x_2$
(4) $\partial L/\partial \lambda_2 = 10 - x_2$

If the derivatives are set equal to zero, and if this system of four equations is solved, the following solutions are found: $x_1 = 10$; $x_2 = 10$; $\lambda_1 = 0$; $\lambda_2 = 1.25$.

Hence, the optimal application of x_1 as well as x_2 is ten units. The shadow price of the first restriction (λ_1 (budget constraint)) is 0 (zero) which means that it would not have been possible to earn more, not even with a higher budget. The shadow price of the other restriction (λ_2) is 1.25 which means that it would have been possible to achieve a profit which would have been MU 1.25 higher if it had been possible to buy one more than the ten units of x_2.

10.3 Examples of Usage

Following this general introduction, some examples of optimisation under restrictions will now be presented.

The budget constraint, as discussed above, is a relatively well-known and well-described restriction. However, restrictions come in different forms and can have different reasons. In the following, we will look at some other possible restrictions.

Presume that the government wants to regulate production either by limiting the production of the product y, or by limiting the application of one or more of the inputs $x_1 \ldots x_n$ (for the sake of simplification only the two inputs x_1 and x_2 will be discussed in the following. This will not constitute any real limitation of the validity of the results when using multiple inputs). Such regulation can be implemented in various ways: (1) A levy can be imposed on the production of y (reduces the price p_y). (2) A tax (levy) can be imposed on one or both inputs (increases the input prices w_1 and w_2). (3) The possibility of buying input can be restricted by limiting access to capital. (4) Production can be limited physically by the introduction of a quota on the production of the product y (e.g. a milk quota). (5) A quota on the application of input (e.g. a requirement to set-aside land) can be introduced. And finally, these measures can be combined.

With the model structures in place, we will now analyse the implications of the various measures. To prepare the analysis, it is presumed that the production function $f(x_1,x_2)$ is *homothetic*,[2] all the marginal products are diminishing, and that

[2] Repeat the homothetic production function in Sect. 4.3.

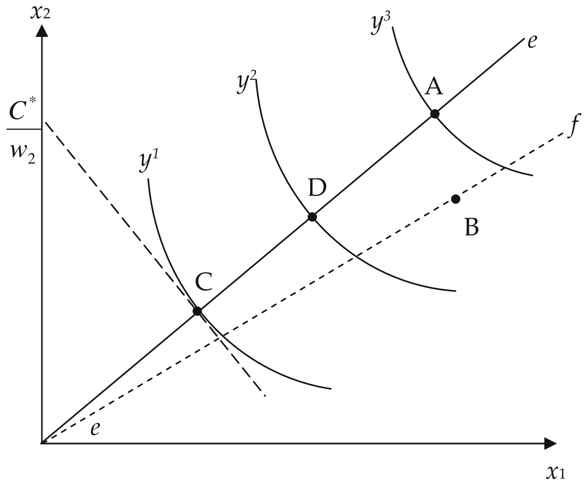


Fig. 10.1 Isoquants and expansion paths

the producer is already maximising profit (point A in Fig. 10.1). If a production function is homothetic, it means that its expansion paths are linear and pass through the zero point, as described in Chap. 4. This assumption simplifies the graphical analysis and does not entail any decisive limitation on the general validity of the results.

1. *The price of y is reduced*: When p_y decreases, the value of the marginal product (VMP_i) also decreases. With given (unchanged) input prices, the fulfilment of the condition for profit maximisation, $P_yMPP_2/w_2 = P_yMPP_1/w_1 = 1$, presupposes higher marginal products at optimum and, thus, a lower production. Hence, adjustment will take place along the expansion path from A towards the south-west (towards D in Fig. 10.1). The production decreases, the application of the two inputs decreases, but the relationship between the two inputs remains unchanged.
2. *The price of input is increased*: When the price of the two inputs is increased with the *same percentage*, the effect would be the same as when the product price decreases: With higher input prices, the fulfilment of the condition for profit maximisation ($P_yMPP_2/w_2 = P_yMPP_1/w_1 = 1$) presupposes higher marginal products at optimum and, hence, a lower production. Hence, adaptation will take place along the expansion path from A towards the south-west. The production decreases, the application of the two inputs decreases, but the relationship between the two inputs remains unchanged. If the price of the two inputs is increased with *different percentages*, the expansion path is shifted. If w_2 e.g. is increased by more than w_1, the expansion path will turn clockwise (e.g. to the expansion path *ef* in Fig. 10.1), and adjustment will take place as a movement from point A onto and down along the new expansion path *ef*, e.g. to point B. The production decreases, the application of the more expensive input x_2 decreases, and the use of x_1 increases compared to the use of x_2.

3. *Limited capital*: If the producer cannot afford to buy all the desired inputs (corresponding to profit maximum), this is really a budget constraint, as described in Chap. 4, Sect. 4.2. If this is the case, production must be reduced. If the producer e.g. only has MU C^* with which to buy input, the adjustment in Fig. 10.1 will take place along the expansion path from A to point C where the corresponding amount of x_1 and x_2 costs precisely MU C^*. The production decreases, the application of the two inputs decreases, but the relationship between the two inputs remains unchanged.
4. *Quota on output*: The production can also be reduced by directly forbidding the producer to produce more than a certain maximum amount (imposing a quota). Presume e.g. that a quota of y^2 kg is imposed, as shown by the corresponding isoquant in Fig. 10.1. If this is the case, the task consists of producing the given amount of y^2 in the cheapest possible way, corresponding to the optimisation task described in Sect. 4.2 above. The adaptation will take place along the expansion path from A to D. The production decreases, the application of the two inputs decreases, but the relationship between the two inputs remains unchanged.
5. *Quota on input*: Imagine a measure whereby the producer is only allowed to use b_1 units of input x_2. This situation is described in Fig. 10.2. Compared to the initial situation (the profit maximisation at point A), the application of x_2 must thus be reduced from b_0 to b_1. But what about the application of x_1? In line with the previous analyses, the adaptation will be expected to take place along the expansion path, so that the new production point will be point D in Fig. 10.2. However, point D is not optimal. There are *no* restrictions on the application of input x_1. And there are no restrictions on the size of the production either. It would therefore be optimal to expand the application of x_1 to the point where the value of the marginal product (VMP_1) is precisely equal to the input price (w_1). This corresponds to point E on the dotted pseudo scale line in Fig. 10.2.

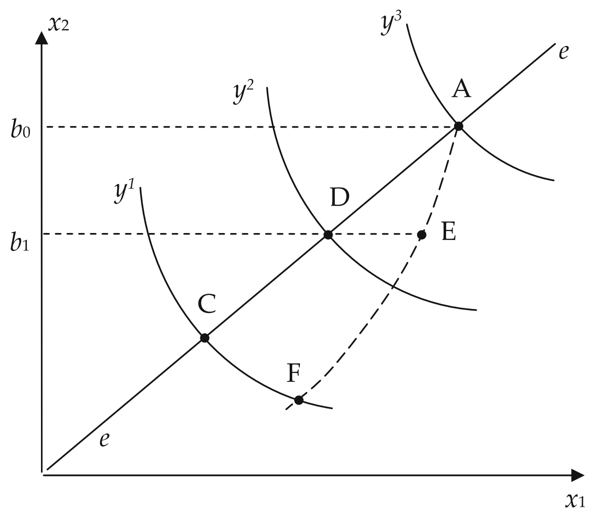


Fig. 10.2 Isoquants and pseudo-scale-line

Hence, point E represents the new optimal production, and the adjustment will thus take place along the pseudo scale line from A to E. The production decreases, the application of the two inputs decreases, and the relationship between the two inputs changes so that more input x_1 per unit of x_2 is used than initially.

The described models can be related to practice as described below.

10.3.1 Set-aside vs. Restriction of Production

Presume e.g. that the regulatory authority wants producers to reduce production of a given output y (e.g. cereal crops) from the present amount of y^3 to y^2 kg (see Fig. 10.2). Land (x_2) and "other input" (x_1) are used in the production. The government is considering whether to use the above option 4 (production quota) or option 5 (input quota). With the first option, a production quota of y^2 kg is imposed on the producer. With the second option, only b_1 hectares of land are permitted to be cultivated, i.e. all remaining land is set aside.[3]

As can be seen in Fig. 10.2, these two "policies" have different effects. Firstly, option 5 (input quota) will not reduce the production all the way to y^2 (the isoquant for point E is placed more to the right than y^2). Secondly, it will result in an increased application of "other input" per hectare of land being cultivated. And this "side effect" might not have been entirely anticipated!

10.3.2 Limited Input

Consider another example that can be of importance in practice. In theory, the assumption is often made that producers maximise profit—i.e. find themselves at point A in Figs. 10.1 and 10.2 before a possible adaptation in connection with regulatory governmental measures. However, this need not be the case in practice. Consideration of risks and (hidden) transaction costs may mean that the producer is initially to be found at another point on the expansion path (or on an entirely different expansion path) than the one corresponding to the profit-maximising production specified from the point of view of the theorist. What would the implication of this be?

Presume e.g. that the producer is initially to be found at point P in Fig. 10.3. The reason why he/she is not to be found at the profit-maximising point A can e.g.

---

[3] The example is naturally somewhat simplified. Hence, the possibility of the available land being used to grow other crops is not considered here. Also, please note that the model in Fig. 10.2 and 10.3 used for land in fact presupposes: (1) the existence of fixed input (e.g. labour) in addition to x_1 and x_2, or (2) that the quality of land declines with the increasing amount of acreage. Otherwise, there would be no profit maximum at A!

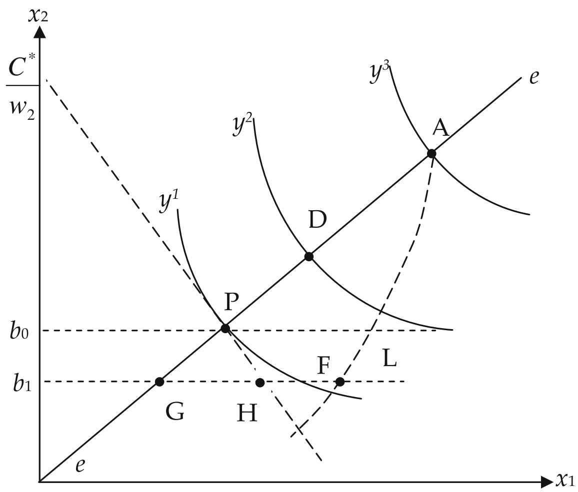


Fig. 10.3 Adjustment under budget constraint

be related to a budget constraint, understood in the way that he/she only has a capital of MU C^* at disposal.

As can be seen in Fig. 10.3, the producer initially uses b^0 units of x_2 (the largest possible production with the given limited amount of capital). It is now assumed that he/she is obliged to reduce the application to e.g. b_1. In line with the previous analysis, this will entail that the producer is expected to adjust production along the dotted pseudo scale line from A: Firstly, adjustment is carried out along the expansion path from the point of reference P to point G. However, here the producer discovers that it will be profitable to increase the amount of x_1, which he/she can now actually afford (as opposed to before). Therefore, x_1 is increased from G towards F which is a point to be found on the pseudo scale line. However, the adjustment does not make it all the way to F as, at H, the total budget of C^* has been used. Hence, in practice, an adaptation from P to H will take place.

Please note that even though production decreases (H is found on a lower isoquant), a relative (compared to x_2), as well as an absolute increase of the application of input x_1 takes place. The reason for this is that, apart from an increase in the marginal product of x_1 when the two inputs are reduced along the expansion line, it is now possible to buy more of x_1—however, not as much as desired, in this example (corresponding to point F).

10.3.3 Compulsory Set-aside of Land

There may be other reasons why the producer is not initially found at the profit maximum point. The initial situation in Fig. 10.3, where only b^0 units of x_2 (land) are used, may e.g. be due to the fact that there is no more land for sale (or lease) in the area. The farmer must therefore be "content" with the b^0 hectares, even though he/she would have liked to buy or lease more land (at "normal" prices).

It is assumed, as previously, that he/she is forced to reduce land to e.g. b_1. This makes land a fixed factor. As shown before, it is therefore not optimal to produce on the expansion path. The optimum is instead found at point F on the dotted pseudo scale line.

The implication of the adaptation seems to be that much more "other input"(x_1) per hectare of land (1 divided by the slope of the line 0F) is now used than initially (1 divided by the slope of the line 0P). However, this is a misinterpretation. Initially, the producer is not found at point P. Land is in fact already part of the initial situation as a fixed input (it was not possible to buy more, even though the producer would have like to!). Therefore, the initial optimal point for the producer was L, and the increase in the application of "other input" per hectare of land, therefore, only corresponded to the change from L to F.

The examples show that the actual adjustment depends on the preconditions assumed to be valid in the initial situation.

10.3.4 Adaptation due to Product Price Decrease

Consider one more example. This could e.g. be the production of finisher pigs with the use of two inputs, fodder (x_1) and stables (x_2). Presume that the producer illustrated in Fig. 10.2 gets access to sufficient capital over time, and that she/he therefore expands production. Both x_1 and x_2 are *variable inputs* in this expansion process, and the adaptation, therefore, takes place along the expansion path until reaching point A.

However, after expansion—unfortunately for our producer—the product price p_y decreases, and so production must—as in case 1 above—be reduced.

As it is, one would expect this reduction to take place along the expansion path—i.e. back along the same way as in case 1. However, the special feature of this example is that the input x_2 (stables) in a situation of reduction, such as the one the producer is now facing, turns into a *fixed input*. This means that the sales price of the previously acquired units of x_2 is now lower than the marginal value at a continued production—even after the huge product price decrease. This situation is illustrated in Fig. 10.4.

The immediate implication of the product price decrease is that the application of the variable input is reduced to point B, where the condition $VMP_1 = w_1$ is assumed to be valid. After this, the production is adjusted over time, corresponding to the adaptation along the new shifted pseudo scale line from point B as the stable facility is worn down. When there e.g. are b_3 units left of the original stable facility, the production will have been reduced to y^2 in point C.

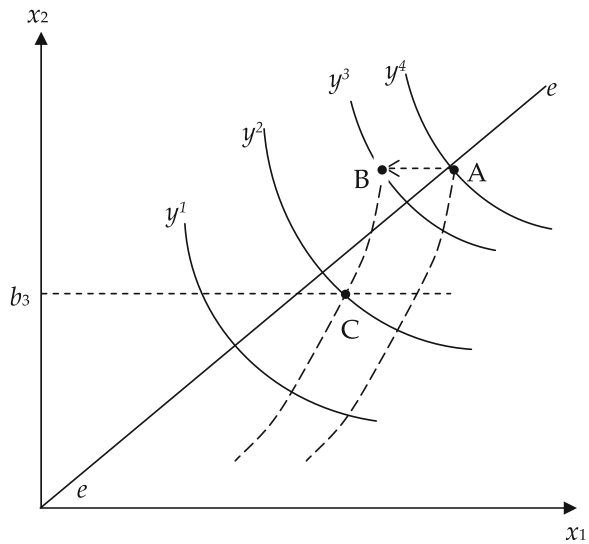


Fig. 10.4 Adjustment due to price decrease

10.3.5 Restriction of Output—Illustrated by Milk Quota Example

The last example of optimisation under a restriction is provided the case when the amount of products that the producer is allowed to produce and sell is restricted.

In 1984, the EU introduced a quota system for the production of milk, the objective of which was to limit milk production in the EU as the price at the time had resulted in overproduction forcing the EU to buy and sell the surplus at considerable cost. The quota system was introduced so that all farmers were only allowed to sell a limited amount of milk at the normal price. If they produced more than this limited amount of milk (quota), a levy corresponding to the normal sales price was be imposed on this amount so that the actual price for the surplus production would be zero MU. This system functioned as a physical restriction on production as there was, of course, no incentive to produce more than the allocated quota.

The problem for the farmers was to decide how to reduce their production of milk. Should it be done by selling some of the cows, or by reducing fodder, which would result in a reduction in each cow's capacity to produce milk. These two approaches can naturally also be combined.

This issue is discussed in greater detail in Rasmussen and Hjortshøj Nielsen (1985), whilst the key conditions in connection with the formulation of, and solution to, the problem are presented here.

The following concepts are used:

x_1: Number of cows
x_2: Amount of fodder (fodder units (FE)) per cow
$y = f(x_2)$: The milk production function for a cow

w_1:	Costs per cow (excluding fodder) (breeding, veterinary services, labour, interest)
w_2:	Price per fodder unit (FE)) of fodder
p_y:	The price of milk
FC:	Fixed costs (buildings and equipment (stable facility etc.))
M:	Milk quota

Using the described conditions, the total product value can be calculated as[4]:

$$TPV = p_y x_1 f(x_2) \tag{10.23}$$

and the total costs as:

$$TC = w_2 x_2 x_1 + w_1 x_1 + FC \tag{10.24}$$

where the first two terms represent the costs related to fodder and the number of cows. These costs will be referred to as production costs (PCs) in the following. Hence, the PCs per cow are:

$$PC/x_1 = w_2 x_2 + w_1 \tag{10.25}$$

The profit π is thus:

$$\pi = p_y x_1 f(x_2) - (w_2 x_2 x_1 + w_1 x_1 + FC) \tag{10.26}$$

The problem can then be formulated as:

$$\underset{x_1, x_2}{\text{Max}}\ \pi \tag{10.27}$$

under the restriction:

$$x_1 f(x_2) = M \tag{10.28}$$

The problem of (10.27) and (10.28) is based on the assumptions that the farmer's present production (before the introduction of the quota system) is optimal and that the introduction of the quota entails that production must be reduced.

To solve the problem, the *Lagrange function* is formulated as:

$$\text{L} = p_y x_1 f(x_2) - (w_2 x_2 x_1 + w_1 x_1 + FC) - \lambda(x_1 f(x_2) - M) \tag{10.29}$$

and L is differentiated with regard to the three variables x_1,x_2, and λ, and the derivatives are set equal to zero resulting in the following conditions for optimality:

$$p_y f(x_2) - w_2 x_2 - w_1 - \lambda f(x_2) = 0 \tag{10.30a}$$

[4] For the sake of simplification, the meat production that is normally part of milk production is disregarded. Please refer to the original article which includes meat production.

$$p_y x_1 MPP_2 - w_2 x_1 - \lambda x_1 MPP_2 = 0 \tag{10.30b}$$

$$x_1 f(x_2) - M = 0 \tag{10.30c}$$

MPP_2 is the marginal product of f with regard to x_2.

Equations (10.30a) and (10.30b) can be written as:

$$-w_2 x_2 - w_1 = (\lambda - p_y) f(x_2) \tag{10.31a}$$

$$-w_2 x_1 = \left(\lambda - p_y\right) x_1 MPP_2 \tag{10.31b}$$

Dividing (10.31a) by (10.31b) and moving around produces:

$$\frac{w_2 x_2 + w_1}{f(x_2)} = \frac{w_2}{MPP_2} \tag{10.32}$$

According to (10.32), the condition for optimal production is that the average PCs per kg of milk (the left hand side in (10.32) are equal to the marginal PCs per kg of milk (the right hand side in (10.32).

That the left hand side in (10.32) represents the average PCs per kg of milk can be seen by dividing the PCs per cow in (10.25) by the total milk production per cow $f(x_2)$ which produces the left hand side in (10.32) as the result.

Differentiating the PCs per cow in (10.25) with regard to the milk production results in the following marginal costs:

$$MC = \frac{\partial(PC/x_1)}{\partial(f(x_2))} = \frac{\partial(PC/x_1)}{\partial x_2} \frac{\partial x_2}{\partial(f(x_2))} = w_2 \frac{1}{MPP_2} = \frac{w_2}{MPP_2} \tag{10.33}$$

which represent the right hand side in (10.32).

As shown above (in Chap. 5), the marginal costs are equal to the average costs at the precise point where the average costs reach the minimum.

The condition for optimal production under quotas stated in equation (10.32) is, thus, that the allocated quota is produced with the lowest possible PCs.

The condition (10.32) can be illustrated graphically, as shown in Fig. 10.5.

The lowest average cost per kg of milk is achieved precisely at the point where the marginal costs are equal to the average costs (*GPC*). This corresponds to a milk production per cow of y_p kg and a corresponding optimal amount of fodder of x_{2p} FE per cow.

If the price of milk is p_{y0} and the optimal production *before* the introduction of the quota system was y_o, then the introduction of the quota system would mean that the amount of fodder for the cows should be reduced from x_{20} FE per cow to x_{2p} FE per cow.

Please note that this only guarantees the fulfilment of the conditions (10.30a) and (10.30b). The fulfilment of condition (10.30c) would entail a production that does not exceed the quota M. If the reduction in fodder from x_{20} FE per cow to x_{2p}

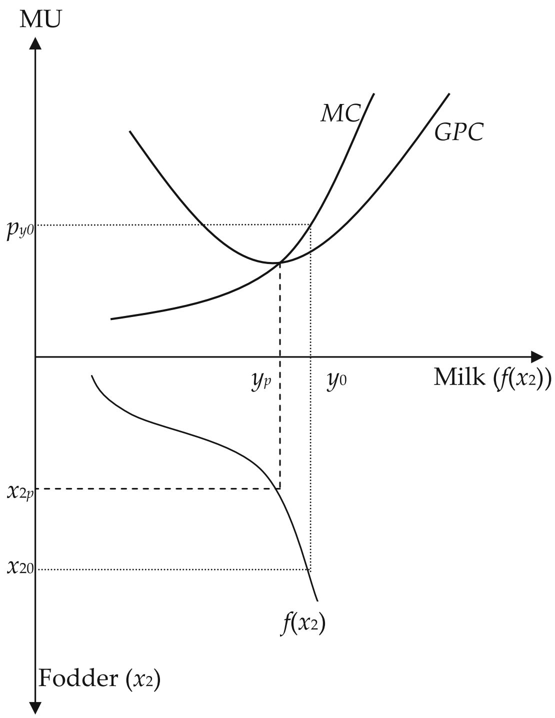


Fig. 10.5 Determination of optimal fodder amount per cow

FE per cow is insufficient, then the number of cows x_1 should also be reduced, so that the condition:

$$x_1 f(x_2) = M \tag{10.34}$$

is also fulfilled.

If (10.34) is in fact fulfilled with the fodder amount of x_{2p} FE per cow, then the number of cows x_1 does not need to be changed, and the adaptation to the quota is carried out exclusively by reducing the amount of fodder from x_{20} FE per cow to x_{2p} FE per cow.

If a reduction in the fodder amount from x_{20} to x_{2p} FE per cow is more than sufficient to reduce the production of milk to the quota, then the present number of cows does not need to be changed. The fodder amount is reduced only to the point where (10.34) is fulfilled (which, as shown here, happens *before* the point with the lowest costs is reached).[5]

Please note that an optimal adaptation may consist of: (1) A reduced amount of fodder. (2) Sale of cows. (3) Both (1) and (2). The shape of the cost curve at

[5] The reason why this is optimal is because a possible further reduction in production will not be motivated by the quota restriction but by a desire to achieve profit maximisation.

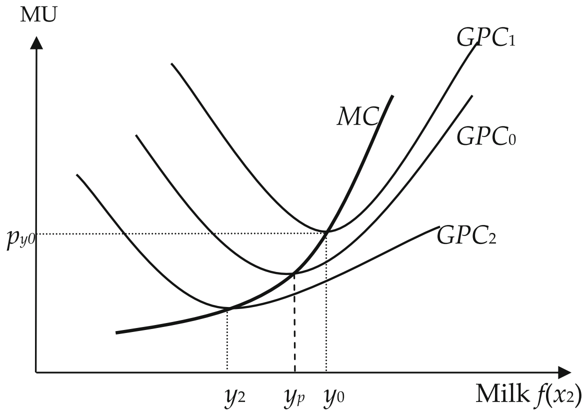


Fig. 10.6 Adjustment of milk production

optimal production before the introduction of the quota will be decisive for the optimal choice.

This is illustrated by a brief analysis. The upper part of Fig. 10.5 has been reproduced in Fig. 10.6.

Initially, when the price is p_{y0}, the PCs are GPC_0, and the optimal production *before* the introduction of the quota system is therefore y_o. The introduction of the quota system results in a reduction of the production of milk per cow to y_p kg.

If the cow costs (w_1) are initially higher, the *GPC* curve would be placed further towards the north-east (see the left hand side in (10.32)). If the average PCs e.g. were GPC_1, the optimal production per cow would remain unchanged y_o, and the adaptation to the quota would be carried out exclusively by selling cows. If, on the other hand, the cow costs (w_1) were lower, so that the PCs e.g. were GPC_2, the optimal adaptation would be carried out by reducing the production of milk (fodder) per cow to y_2, and only then would cows be sold—if necessary.

The cow costs (w_1) are, thus, crucial for the way in which the adaptation should be carried out. If the cows are relatively expensive to maintain (w_1 is high), the adaptation will primarily be carried out by selling cows. If, on the other hand, the cow costs are relatively low, adaptation will primarily be carried out by a reduction in the amount of fodder.

In the long run, the fixed costs (*FC*) (stable facility etc.) become variable costs as, in the long run, the stable capacity can be adjusted when making new investments. If this is the case, the original fixed costs (*FC*) are changed to cow costs (w_1), and the composition of the total costs (TC) in (10.24) is changed correspondingly. This means that, in the long run, w_1 increases and the *GPC* curve in Fig. 10.6 is thus shifted towards the north-east, and the probability of an adaptation based on the sale of cows is increased.

Hence, it can be concluded that the optimal adaptation strategy differs in the short and the long run. In the short run, when an important part of the costs of keeping cows are fixed, it will primarily be an issue of adaptation by reduction in the amount of fodder. In the long run, when all costs are variable, adaptation will primarily be carried out by selling cows.

It is up to the reader to demonstrate that if equilibrium is to be achieved in the long run—with the milk price providing full cost coverage—adaptation to a milk quota, in the long run, will consist solely of the sale of cows and an unchanged supply of fodder for the remaining cows.

References

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

Rasmussen, S., & Hjortshøj Nielsen, A. (1985). The impact of quotas on the optimal adjustment of milk production at the farm level. *European Review of Agricultural Economics, 12*, 351–364.

Chapter 11
Economies of Scale and Size

11.1 Introduction

The two concepts *economies of scale* and *economies of size* describe what happens to production or costs when the size of the firm changes (increases). Economies of scale describe how much production increases when the firm increases its scale of production, i.e. increases all (both fixed and variable) inputs by a common proportionality factor. Economies of size describe what happens to cost per unit of output when production increases in a cost minimising way.

Empirical observations show that larger companies are often more productive than smaller companies, and that large companies therefore produce at lower unit costs than small companies. This advantage seems to drive the structural development of industries through time, so that companies typically become larger and larger. However, history has also shown that there may be a limit to how large a company can grow and still reduce unit cost. The very large state or collective agricultural companies under the former communist regime in the Eastern European countries are often used as examples of companies which, from an economic point of view, were probably too large, and one could therefore talk about the disadvantages of companies being too large.

This chapter starts out with a formal definition of the two concepts, economies of scale and economies of size, and related concepts. The relationship between the size of the production and costs has already been discussed in Chap. 5. However, in Chap. 5 the focus was primarily on the relationship between costs and production within the framework of a *given fixed asset*. In the present chapter we allow the firm to increase production by investing in new and larger assets, and we extend the discussion of the relation between short and long run costs, as initiated in Sect. 5.3. The long run average cost curve derived in Chap. 5 is an essential tool for analysing and describing the economies of size in the following.

Economies of scale and economies of size are closely related concepts, and many textbook authors and economists in general do not distinguish between the two. However, in this book we will make a distinction, because it is confusing to

S. Rasmussen, *Production Economics*, DOI: 10.1007/978-3-642-14610-7_11,

mix the two as they each have their own specific meaning: Economies of scale is a *technical term* that describes the properties of the production function. Economies of size is an *economic term* that describes the behaviour of the (long run) cost function.[1]

11.2 Economies of Size

The cost curve for a company was derived in Chap. 5 (see Figs. 5.2 and 5.3) above. The specified cost curve has been derived based on the underlying precondition that production is carried out within the framework of a given *fixed asset* with associated given *fixed costs* (*FC*). Within the framework of such given assets, the average total costs will have a shape corresponding to the *ATC* curve in Fig. 5.3, i.e. the average costs will first decrease with increasing production and then increase with increasing production.

This cost curve (Figs. 5.2 and 5.3) is per definition a *short run cost curve*. The short run definition is related to the fact that parts of the input factors are fixed factors.

In Chap. 9, the derivation of the optimal production for the company on the basis of this cost curve was illustrated (see Fig. 9.1). The company's supply curve was shown to be equal to the marginal cost curve. As this concerns the short run, it is a matter of *short run marginal costs* and hence of a *short run supply curve*.

In the long run, it is possible to adjust the fixed asset. This would either consist of reducing the size of the plant through re-investment in a smaller plant when the old plant is worn down, or expanding the plant through new investment in a larger and possibly better plant.

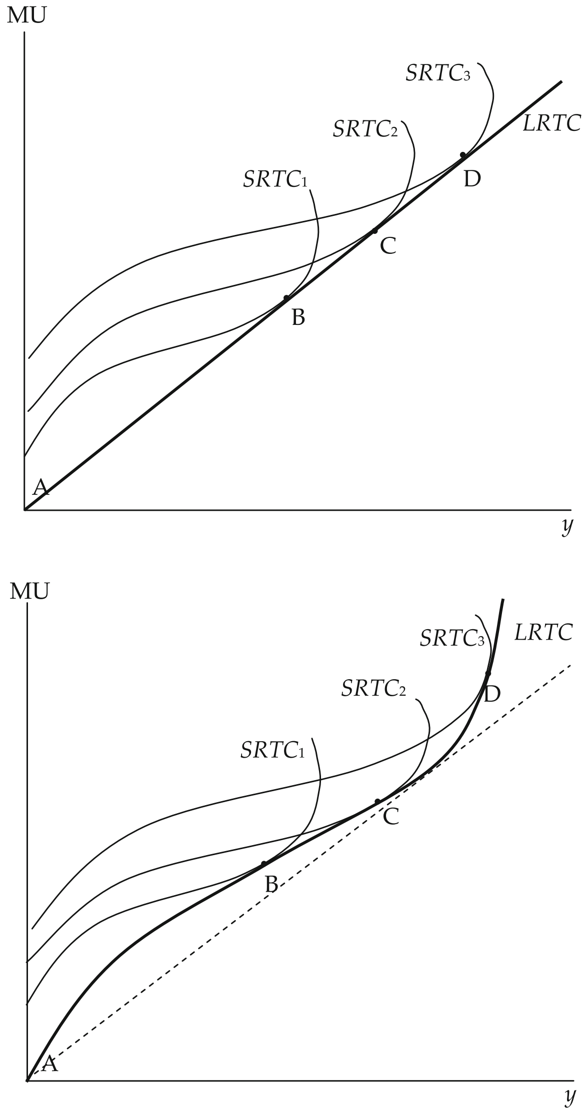
The various possibilities facing the company to vary production in the long run can be described in various cost curves, as shown in Fig. 11.1.

The figure only shows three short run cost curves, corresponding to three different plant sizes. These three curves are now referred to as *SRTC* to indicate that they are *Short Run Total Cost* curves.

If you imagine the possibility, in the long run, that a plant could have any size (the short run curves are placed closely together), i.e. that any plant size can be chosen in the long run, then the curve for the *long run total costs* could be drawn as an *envelope curve* for the short run curves which are illustrated by the curve with the bold line in Fig. 11.1. The upper part of the figure shows a technology that, in the long run, results in a linear increasing long run cost curve. The bottom part of the figure shows a technology for which the long run costs are first digressively increasing (until after point C) and then progressively increasing.

[1] In this book we distinguish between economies of scale and economies of size in the same way as Chambers (1988), Debertin (1986) and other (agricultural) economists. Notice that general microeconomic or managerial textbooks such as Gravelle and Rees (2004), Maurice and Thomas (2002), etc. typically use scale economies to describe both concepts.

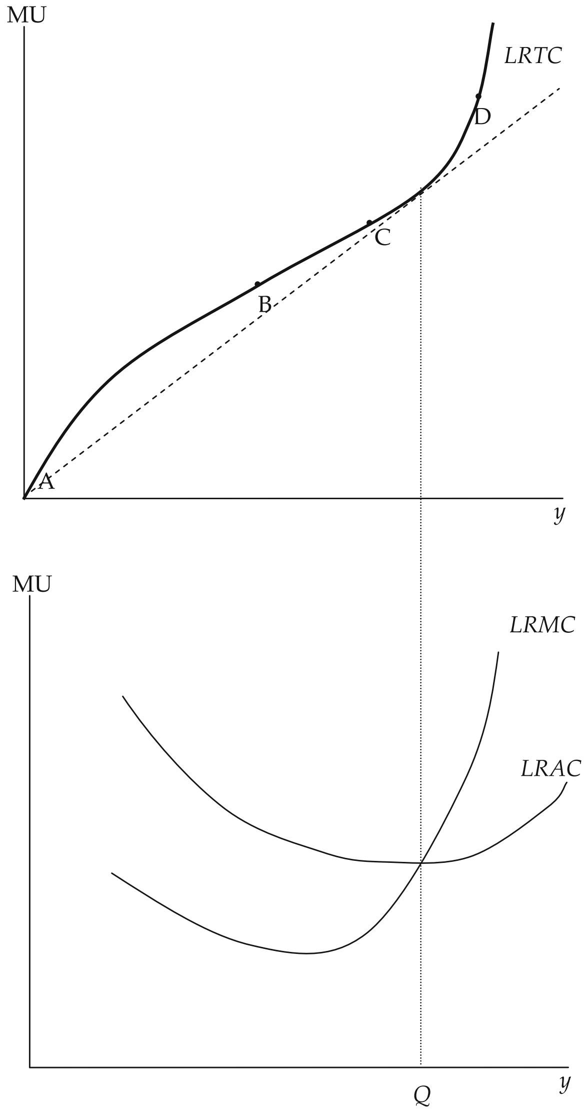
Fig. 11.1 Deduction of long run cost curves



In the upper part of the following Fig. 11.2, the bottom parts of the *LRTC* curves from Fig. 11.1 are repeated. In the bottom part of Fig. 11.2, the long run average costs (*LRAC*) and the long run marginal costs (*LRMC*) have been drawn. As can be seen, the minimum of the long run average costs is found at a production of *Q* units.

Economies of size is now defined as existing as long as the long run average costs (*LRAC*) are decreasing (i.e. to the left of point *Q*). Similarly, *diseconomies of size* is defined as existing as long as the long run average costs (*LRAC*) are increasing (i.e. to the right of point *Q*).

Fig. 11.2 Deduction of long run cost curves



To summarise, the term *economies of size* is used to describe a situation in which the total cost per unit of output decreases as the firm expands its output. The term *diseconomies of size* is used to describe a situation in which the total cost per unit of output increases as the firm expands its output.

There are a number of possible *reasons for companies facing economies of size*. Better utilisation of existing capacity will often be of key importance. Owning a machine that can handle a larger production will provide the opportunity of distributing the fixed costs to more product units and thereby achieve lower unit costs. If there is any idle capacity, this idle capacity might be combined with the new investment capacity to achieve an advantage. Such *synergies* can be of huge importance and managers who are good at identifying such available resources in the company and combine them with new assets will often be able to achieve cost advantages.

Other advantages can be found on the price side. At large production volumes, it might be possible to achieve lower input prices due to quantity discounts. Such a quantity discount can be "real", in the way that the supplier saves expenses by e.g., delivering raw materials or intermediate inputs in bulk instead of in bags, which benefits the buyer. It could also be that the company, due to its size, achieves a stronger position in the market, a position that can be used to pressure suppliers to deliver at lower prices.

Finally, the technical development should be mentioned as this entails a continuous development of new assets (machinery, buildings, and fixtures) that are better and more cost efficient than the old assets. There also seems to be a tendency for new technology to favour large scale production in particular, so that it is precisely through large scale production that the new technology is most (cost) beneficial. Hence, the technological development can be illustrated by the new cost curves which provides for lower costs, especially in connection with increasing product quantities.

The reasons for *diseconomies of size* can e.g. be related to situations where the asset/company becomes so big that the operational manager no longer has an overview of it all, and things therefore start to go wrong. There can be processes that are not initiated on time, or there might not be time to monitor production. The classical example from agriculture is when the acreage becomes so large that the activities on the outermost field cannot be monitored. In general, increasing production without a corresponding increase in management resources is a typical reason for diseconomies of size.

11.3 The Optimal Size of the Firm

The long run average cost curve derived in Figs. 11.1 and 11.2 is a planning curve for the firm, and it shows how it is possible, in the long run, to adjust the unit cost by adjusting the capacity of the firm. As long as economies of size exist, i.e. as long as the average cost decreases, a rational company owner, being a price taker in the output market, should expand the plant and the production until the point where economies of size are fully exploited, i.e. at point Q in Fig. 11.2. The *optimal size* [measured in units of output (y)] of a company is therefore a company which produces the amount Q and with a production capacity corresponding to the plant with the lowest short (and long run) average costs at this production level. Using Fig. 11.1 above, the optimal production capacity in this example is a plant with a capacity between plants 2 and 3.

It is important to differentiate between optimising production within the framework of a certain capacity (plant), and long run optimisation. In the first case, the tool for optimising production is given in Sect. 3.3 in Chap. 3 and Chap. 9. In the second case, as described in this chapter, it is the long run average cost curve.

But how is it possible—to talk about an *optimal* size (production) of the firm without knowing the product price, as we do in this second case?

The reason is that in the long run, the product price under perfect competition will adjust according to the long run average costs, and the long run product price is precisely equal to the lowest long run average costs (Repeat Sect. 9.3). In connection with a long run strategy for the expansion of the firm, it will therefore be rational to plan an expansion of the firm so that it corresponds to the firm size and the production *Q* in Fig. 11.2.

If, as described, an optimal size of the plant/firm/company exists, then why do not all companies within an industry have this particular size? For example, within agriculture, why are not all farms the same size?

There are several reasons for this. Firstly, the possibility for investing in new and bigger plants differs from firm to firm. There may be budget constraints which mean that there is no money to make the necessary investment to expand the plant. Maybe the necessary inputs are not available. If the expansion e.g. requires the purchase of more land, such expansion presupposes that there is land for sale in the area. There might also be local restrictions in the form of legislation and regulations limiting expansion possibilities.

Secondly, both the short run as well as the long run cost curves can vary from one company to another. What might be an optimal size of the firm for one producer might not be an optimal size at all for other producers. This can be due to differences in both physical production relationships (the production function) as well as in input prices (measured as opportunity costs!).

11.4 Economies of Scale

Contrary to the concept economies of *size*, economies of *scale* is a purely technical concept. The term *economies of scale* refers to what happens to the amount of output if *all inputs are increased proportionally*. If output increases *more than* inputs, i.e. *if all* inputs are increased by for instance 5%, and the production increases by 6%, then there are *increasing returns to scale*. If output increases less than inputs, i.e. if all inputs are increased by 5% and the production increases by 4%, then there are *decreasing returns to scale*. And finally, if output increases in the same proportion as (all) inputs, i.e. if all inputs are increased by 5% and the production increases by 5%, then there are *constant returns to scale*. This is a property of the production function, and it is not an economic term as such.

The formal definition of returns to scale is the following: if the product *y* is produced using *p* different inputs (including both variable *and* fixed inputs) and the production function is *f*, then:

$$y = f(x_1, x_2, \ldots, x_p) \tag{11.1}$$

If all p inputs are increased by multiplying each input by a factor t ($t > 1$), and if the output y also increases by the same factor t, i.e.:

$$tf(x_1, x_2, \ldots, x_p) = f(tx_1, tx_2, \ldots, tx_p) \tag{11.2}$$

then the production function exhibits constant returns to scale. If output y increases by less than the factor t, such that:

$$t^n f(x_1, x_2, \ldots, x_p) = f(tx_1, tx_2, \ldots, tx_p), \quad (n < 1) \tag{11.3}$$

then the production function exhibits decreasing returns to scale. Finally, if output y increases by more than the factor t, such that:

$$t^n f\left(x_1, x_2, \ldots, x_p\right) = f\left(tx_1, tx_2, \ldots, tx_p\right), \quad (n > 1) \tag{11.4}$$

then the production function exhibits increasing returns to scale.

Notice that economies of scale is normally interpreted as a local measure, i.e. the returns to scale may change from one part of the production function to the other. An example of a one input production function ($p = 1$) is shown in Fig. 11.3. To the left of point A this production function has increasing returns to scale, at point A there are constant returns to scale, and to the right of point A there are decreasing returns to scale.

For some production functions, the returns to scale is the same over the total domain. In this case we say that the specific returns to scale applies globally. The Cobb–Douglas production function is one such production function which has either constant, increasing or decreasing returns to scale globally. To illustrate, consider the following Cobb–Douglas production function with p inputs, which includes all inputs used in the production, i.e. also the inputs which are normally considered fixed in the short run (land, buildings, etc.):

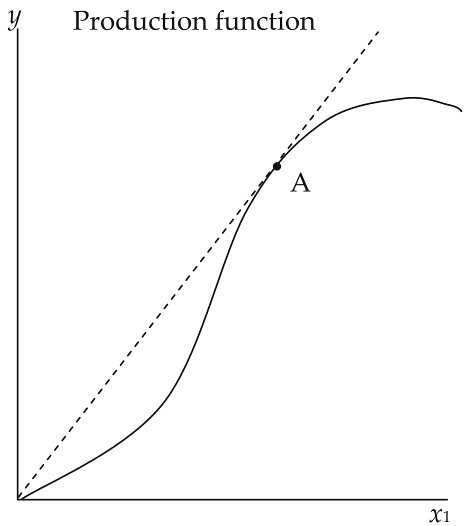


Fig. 11.3 Varying returns to scale

$$y = f(x_1, x_2, \ldots, x_p) = Ax_1^{b_1}x_2^{b_2}, \ldots, x_p^{b_p} \quad (11.5)$$

If we multiply the input vector by a factor t, then the production y is:

$$y = t^{(b_1+b_2+\cdots+b_p)}Ax_1^{b_1}x_2^{b_2}\cdots x_p^{b_p} = t^n f(x_1, x_2, \ldots, x_p). \quad (11.6)$$

This means that if the sum of the parameters ($b_1 + b + \cdots + b_p$) is less than 1, then the Cobb–Douglas production function has decreasing returns to scale ($n < 1$); if the sum is greater than 1, then it has increasing returns to scale; and if the sum is equal to 1, it has constant returns to scale.

From an empirical point of view, it seems reasonable to assume that if all inputs are increased by e.g. 50%, then production would also increase by 50%. The motivation is that the production can always be scaled up: If a farmer is already running two (completely identical) farms, and then buys one more completely identical farm (all inputs are increased by 50%), then one would expect that the production would also increase by 50%.

The problem with the Cobb–Douglas production function is that while the assumption of constant returns to scale could be true locally (as here with the purchase of farm no. 3), it is far from certain that the assumption is true globally. What happens e.g. to the production if the farmer, who used to have two farms, sells one farm and, furthermore, reduces the size of the last one by half? To what extent would that affect production? Would it be reduced by 25% compared to the original level? If, in this situation, production is not reduced to 25%, then the preconditions for using the Cobb–Douglas function with constant returns to scale for a (global) description of the production will not be present in practice.

The problem is that, in practice, it is difficult to imagine situations in which all inputs are in fact increased by the same percentage. Often part of the production factors are fixed inputs that do not change. Consider for instance the production factor "Management" in the example above. If the same farmer, who previously managed two farms, now manages three farms, then the production factor "management" does not increase by the same factor as the other inputs. Imagine, as another example, a farmer who builds a new pig stable for his animals without also buying more farm land or expanding machinery correspondingly. Both examples are not unusual and demonstrate the general problem of using the returns to scale concept in relation to planning and the optimisation of production. In practice, it is unusual that production is expanded by increasing all inputs proportionally. Hence, the concept returns to scale as presented here, seems to be more of a theoretical concept rather than a practical management tool for planning. The relevant question is, therefore, why is the concept at all relevant? Why not just focus on the *size* concept, as presented in Sect. 11.2, and leave the scale concept for purely theoretical exercises?

There are two reasons. First of all, economies of scale is an informative way to describe the (local) properties of a production function. Secondly, the returns to scale concept is appropriate to use in order to *identify* economies of size, and to identify the long run optimal size of the firm under certain conditions.

To see why, consider the dual relationship between the production function and the cost function derived in Chap. 5. As shown in Chap. 5, the cost function is a mirror image of the production function (see Fig. 5.1). In a similar way, it is possible to derive the production function as a mirror image of the cost function.[2] Therefore, there is a dual relationship between the long run total cost function (*LRTC*) in Fig. 11.2, and the long run production function in Eq. (11.1). If the production function is homothetic (with a linear expansion path through the origin), then there is a one-to-one dual relationship between the cost function and the production function, and *increasing returns to scale* is equivalent to economies of size, decreasing returns to scale is equivalent to diseconomies of size, and the optimal size of the firm is where the production function has constant returns to scale. If the production function is not homothetic, the relationship between returns to scale and size economies is unclear. However, in any case where the long run production function has increasing returns to scale then there is also economies of size (decreasing average total costs), and it would therefore be profitable to expand production. The returns to scale concept may, therefore, be used to identify the lower limit of the (economic) optimal firm size. Accordingly, the firm size at which the returns to scale is one is called the *technical* optimal firm size.

11.5 Summary

To summarise the discussion of the two concepts *scale* and *size*, consider the illustration in Fig. 11.4.

Figure 11.4 presents a two input diagram with a number of isoquants. The numbers on the isoquants represent the quantity of output produced. The firm produces at point *A*, which is a point on the expansion path *KK*, where the production is 3 units of y. At this point, there is *increasing returns to scale* because the isoquants become closer and closer moving outwards along the scale line *LL*. After 5 units of output there is decreasing returns to scale, as the isoquants become wider and wider apart moving outwards along the scale line *LL*.

The firm minimises costs as it produces the 3 units on the expansion path *KK* at point *A*. If the firm considered increasing its size, then it should increase production along the expansion path *KK*, which would ensure cost minimisation.

Notice that returns to scale is measured along the scale line *LL*, while economies of size is measured along the expansion path *KK*. In the special case that the production function is homothetic, the expansion path would be a straight line through the origin, and the expansion path would be identical to the scale line. Thus, if a production function is homothetic, production economic analyses are simplified.

---

2 See the conditions and theoretical explanation in Chambers (1988).

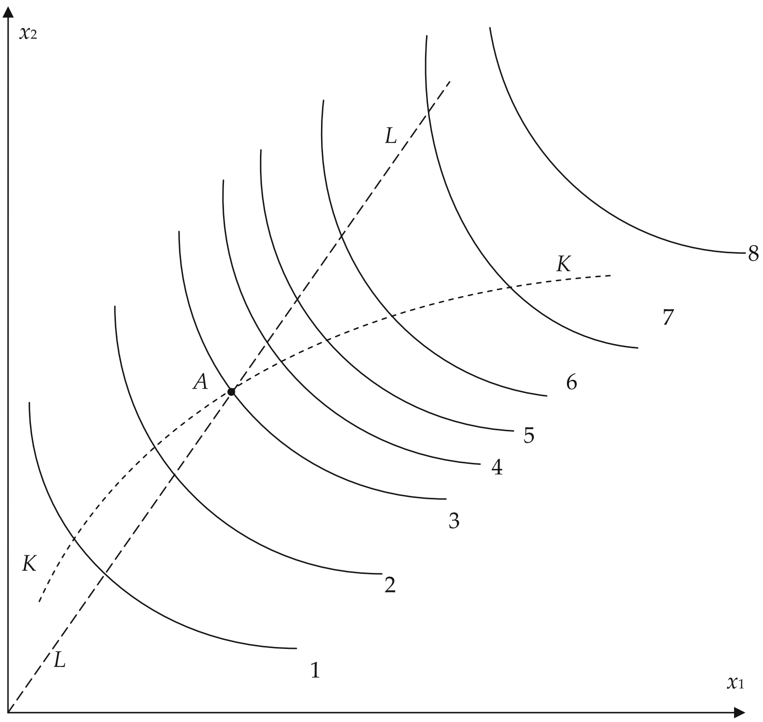


Fig. 11.4 Isoquants, scale line (*LL*) and expansion path (*KK*)

References

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

Debertin, D. L. (1986). *Agricultural production economics*. New York: Macmillan Publishing Company.

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

Maurice, S. C., & Thomas, C. R. (2002). *Managerial economics* (7th ed.). Singapore: McGraw-Hill.

Chapter 12
The Fixation of the Production Factors

12.1 Introduction

Fixed production factors have previously been defined as being production factors *the quantity of which cannot vary within the planning horizon under consideration*. Hence, they are characterised by the fact that the company is unable/unwilling to buy or sell production factors.

However, how are these fixed production factors "generated"? What is the reason why some production factors are fixed whilst others are variable? For how long is a production factor actually fixed? And what is the financial implication of some production factors being fixed? As such, fixed production factors are supposed to entail restrictions which we would like to avoid.

It is questions like these that will be discussed in this chapter. In this connection, there will be a particular focus on the description of the ways in which the fixation of production factors affects the economic behaviour of the producers.

12.2 Degree of "Fixation"

Production factors can be considered "fixed" for one or several of the following reasons:

(1) The production factor *cannot be sold.*
(2) The production factor *cannot be purchased.*
(3) There is *no wish to sell* the production factor.
(4) There is *no wish to purchase* the production factor.

Items 1 and 2 are purely *physical restrictions*. Regarding item 1, these could be production factors which—regardless of the willingness and the price—cannot physically be sold. An example from farming could be a slurry tank for keeping

S. Rasmussen, *Production Economics*, DOI: 10.1007/978-3-642-14610-7_12,

livestock manure. Imagine a slurry tank cast in cement and partly buried in the ground. Such a slurry tank cannot be physically moved and therefore—unless the entire farm is up for sale—it is referred to as a fixed production factor. Regarding item 2, this could be production factors which—regardless of the willingness and the price—cannot physically be purchased. An example could be a building which it would take at least 1 year to build. When making plans for the year to come, the production factor, *building*, will be a fixed factor according to the definition in item 2. (However, if the period of time is longer, e.g. a 5 year planning horizon, it will be a variable production factor as there will be time for erecting the building.)

Items 3 and 4, on the other hand, are purely *financial restrictions*. Item 3 indicates that, even though it would be possible to sell, this is not an option as there is no incentive because the price is too low. Item 4 indicates that, even though it would be physically possible to purchase, this is not an option as there is no incentive because the price is too high.

Items 3 and 4 can also be attributed to purely *behavioural conditions* related to planning. A purchase or sale might simply not be part of the planner's agenda, and even though this is both a technical and financial possibility, the planner is unwilling to consider this option (yet). If this is the case, the situation could be ranked alongside items 1 and 2.

In the following, we will primarily be discussing the financially determined fixation. However, this does not mean that we will not look at fixation that is due to physical restrictions or behavioural conditions. To base the calculation on an infinitely high purchase price and a sales price of zero (or less) would have the same effect, and such purchase and sales prices can simply be introduced in the following economic models to see how such behavioural conditions are illustrated.

12.3 Theory of Fixation of Production Factors

According to the production economic theory discussed in the above chapters, a profit-maximising company—being the price taker in the product market—will add the input factor x_i to the production of the product y in the period t in an amount such that the value of the marginal product (VMP_{it}) (the product value of the last added unit) is equal to the marginal factor costs (MFC_{it}) (see (7.30) in Chap. 7), i.e. so that:

$$\mathrm{VMP}_{it} = \mathrm{MFC}_{it} \tag{12.1}$$

The marginal factor costs are in the simplest example equal to the purchase price of input (w_i). Regarding a more general analysis of the fixation of production factors, the marginal factor costs are more generally defined as the incremental costs related to the use of the last unit of the production factor in question. This incremental cost is calculated according to the opportunity cost principle according to which cost is defined as the revenue forgone by not using the production factor for the best alternative usage.

Regarding resources (production factors) that are *not* (yet) owned by the company, the marginal factor costs of purchasing and using the resource in a period t will be equal to the price of a purchase at the beginning of the period (V_{it}) minus the sales price at the end of the period (V_{it+1}) (depreciation). Added to this are the current operating costs (fuel, repair etc.) (C_{it}). Finally, an alternative return on the capital should be considered, as the MU V_{it} that was used in the purchase could alternatively have been deposited in the bank and have yielded an interest of rV_{it}, where r is the market interest rate (the calculation interest rate). Hence, if depreciation is described as $\Delta V_{it} = V_{it} - V_{it+1}$, the marginal factor costs can be calculated as:

$$\mathrm{MFC}_{it} = C_{it} + \Delta V_{it} + rV_{it} \tag{12.2}$$

Regarding the production factors that are already owned by the company, the marginal factor costs are, in principle, calculated in the same way.[1] The depreciation ΔV_{it} is calculated as $\Delta V_{it} = V_{it} - V_{it+1}$. The decisive difference compared to the purchase situation is, however, that the value of the asset at the beginning and end of the period is calculated as the potential sales value at realisation (sale). If the asset, therefore, is of a nature that results in the potential sales value at realisation being zero, then V_{it} and ΔV_{it} will be equal to zero, and MFC_{it} in the Formula (12.2) will only include C_{it}. This is e.g. the case with the slurry tank mentioned earlier.

According to Eq. 12.1, it will be optimal to acquire (purchase) more of the input factor in question if the value of the marginal product is higher than the marginal factor costs related to a purchase. If the value of the marginal product is lower than the marginal factor costs at sale, it would be optimal to sell units of the production factor in question. Finally, if the value of the marginal product is between the marginal factor costs of a purchase and the marginal factor costs of a sale, it would neither be optimal to purchase or sell units of the production factor. *If this is the case, the production factor is a fixed input based on purely economic criteria.*

If we describe the marginal factor costs of a purchase as $\mathrm{MFC}_{it}(k)$ and the marginal factor costs of a sale as $\mathrm{MFC}_{it}(s)$, the following criteria can be formulated:

$$\begin{aligned} \mathrm{VMP}_{it} > \mathrm{MFC}_{it}(k) &\Rightarrow \text{Purchase more input (variable)} \\ \mathrm{VMP}_{it} < \mathrm{MFC}_{it}(s) &\Rightarrow \text{Sell (some of) the input (variable)} \\ \mathrm{MFC}_{it}(s) < \mathrm{VMP}_{it} < \mathrm{MFC}_{it}(k) &\Rightarrow \text{No purchase or sale (fixed input)} \end{aligned}$$

[1] The marginal factor costs of production factors that are owned by the company represent the income forgone from not renting out the production factor in question (e.g. to the neighbour). With regard to the opportunity cost principle, the income forgone in connection with the best alternative usage should be used.

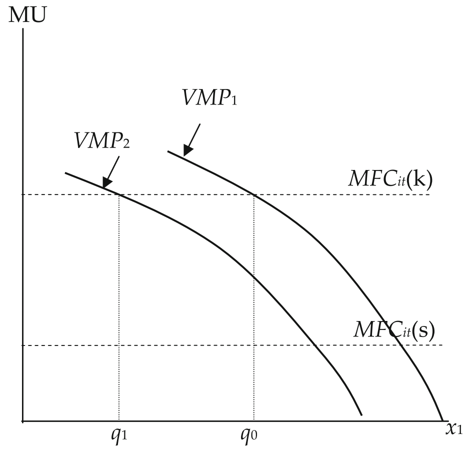


Fig. 12.1 Illustration of fixed inputs

Hence, fixed input can be defined as being those of the company's production factors that have a use value (also called quasi rent) (VMP_{it}) between the marginal factor cost of purchase and the marginal factor cost of sale.

The discussed relationships are illustrated graphically in Fig. 12.1. As a point of reference, it is presumed that the marginal factor costs of purchase of the input in question (e.g. a machine) is $\text{MFC}_{it}(k)$ and the value of the marginal product (the machines *use value* in the company) corresponds to the curve VMP_1. According to the criterion for profit maximisation, it is optimal to buy q_0 units of the production factor.

Presume now that, after the investment in the q_0 units, the price of the product y falls and that the value of the marginal product thus decreases to the curve VMP_2. Under "normal" circumstances, the application of input would be expected to be reduced to q_1 units where the new VMP curve intersects the curve of the marginal factor costs $\text{MFC}_{it}(k)$.

However, this is not optimal. When selling the previously purchased production factor, the example here produces a very low sales price, and if the calculation is based on the actual opportunity costs, the marginal factor cost of the sale is only MU $\text{MFC}_{it}(s)$.

With this low factor cost, it is *not* optimal to reduce the amount of input. With the original amount q_0, VMP_2 is much higher than the marginal factor costs of a sale and it is therefore not profitable to sell, as the asset generates a positive rent. The use value is higher than the (opportunity) costs! On the other hand, it is not profitable to buy more input either. The marginal factor costs of purchase are considerably higher than the use value of VMP_2.

The implication is that, after the purchase of the q_0 units, the production factor has become a fixed input. It is neither profitable to purchase nor to sell units, even though the product price changes, i.e. even though VMP_2 moves up or down. Only in the case that the price falls so much that VMP becomes lower than $\text{MFC}_{it}(s)$ at the amount q_0, would it be profitable to sell some of the input and, thus, reduce production capacity.

The theory about fixed factors discussed here can be used to explain why producers often do not react to price changes, as would be expected. When the product price decreases (the VMP curve drops), the optimal supply of input would be expected to be reduced. However, when—as shown here—there can be huge differences between the marginal factor costs of a purchase and sale, even considerable price decreases may still not affect production. Farmers who have invested in pig stables and other production equipment—expecting continued high prices for pork—will not reduce their production due to a subsequent fall in pork prices. The reason for this is that cost savings of a reduction are limited, as the marginal factor cost in connection with a realisation of (part of) the asset is relatively low, or maybe even zero.

12.4 Disadvantages (and Advantages) of Fixed Factors

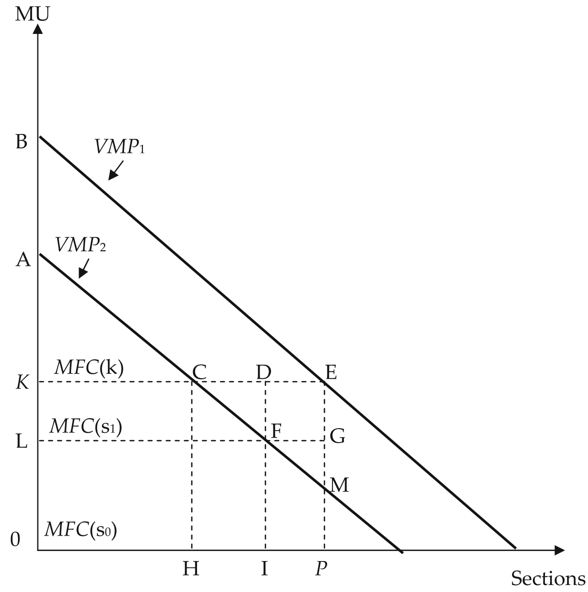
As it is, fixed production factors are generated in situations in which there is a difference between the marginal factor costs of a purchase and sale. The greater the difference, the greater the probability of the factor being a fixed input factor. The borderline case is a situation in which the purchase price is infinitely high and the sales price is zero. Such factors will always be fixed factors.

The question is whether it is a disadvantage for the producer that the production factors become fixed. Generally, fixed factors must be expected to be a disadvantage, and should therefore be avoided.

To illustrate this, presume that a farmer has established a production asset, e.g. a pig stable. The investment was based on the expectation that the price of pork would correspond to the curve VPM_1 in Fig. 12.2. The marginal factor costs in connection with the investment are—for the sake of simplification—assumed to be constant. Let us imagine that the pig stable has been built as a complex of many small sections, and that each of these many sections cost the same, no matter how many sections are built. The capital for the investment has all been borrowed from the bank, and it has been agreed with the bank that the loan is repaid with interest over the expected lifetime of the stable facility with a fixed amount of interest and instalments of MU K per year per section. On the basis of the preconditions mentioned, the farmer builds a pig stable which, according to the criterion for profit maximisation, is the optimal size, consisting of a total of P sections (see Fig. 12.2).

If the preconditions (the price expectations) are met, the farmer will achieve a profit (calculated according to the accounting principle) corresponding to the triangle area KBE in Fig. 12.2. This corresponds to the difference between the gross profit (profit before deduction of capital costs for the stable facility) (the area 0BEP) and the costs (the area 0KEP).

Fig. 12.2 Costs of fixation



After the investment, the product price decreases so that the value of the marginal product is equal to VPM_2. The question is, how the degree of fixation will influence profit.

Example 12.1 If the marginal factor cost is the same for purchase and sale ($MFC(k) = MFC(s) = K$), then this can be referred to as a completely variable input factor. Each individual section can be sold at the same price for which they were originally purchased.[2] And the revenue from the sale is used to reduce the bank loan correspondingly. The adaptation is carried out by selling *P* minus *H* sections, and the costs are thus reduced correspondingly. The profit after the price decrease is thus equal to the area *K*AC, and *the loss is thus equal toABEC*.

Example 12.2 If the marginal factor cost calculated at the sale is lower than the marginal factor cost at the purchase, though not zero, this is referred to as a *conditional fixed factor,* as the fixation of the factor is *conditioned* by how much the product price falls. In the example in Fig. 12.2, it is presumed that the marginal factor cost calculated at a sale is $MFC(s_1)$. As a result of the price decrease, it will be optimal to sell a smaller part of the asset, namely *P* minus *I* sections. The cost saving IFG*P* is used to reduce the bank loan, and the costs are thus reduced correspondingly. The loss that resulted from the price decrease is higher than before as *the loss will be ABEC plus CEGF*.

[2] In practice, this is seldom done, even for production factors that are easily recognised as being completely variable. Normally, so-called *transaction costs* would be involved which reflect the difference between the purchase and the sales price.

Example 12.3 The last example presupposes that the marginal factor cost at a sale is equal to zero (the alternative value of the stable is zero). This means that the marginal factor costs calculated at the sale follow the horizontal axis in the coordinate system in Fig. 12.2 (MFC(s_0)). Under these conditions, it is *not* profitable to sell the stable sections as the use value of the stable after the price decrease (VPM_2) is still higher than the marginal factor cost at a sale. The production therefore continues unchanged in *P* stable sections. *The loss that results from the price decrease is calculated as ABEC plus CEGF plus FGM.*

The example shows that a fixed production factor results in a loss. This loss increases with the degree of fixation, as the biggest loss occurs when the alternative value of the asset is zero (the marginal factor cost at a sale is equal to zero).

But are there no advantages in connection with fixed factors? Yes, there are. If the production in connection with a product price decrease is reduced, as described in Example 12.1, in the case of a completely variable production factor there would be a need for a new investment if the product price increases again. If the new investment takes time—compared to a situation in which the fixed asset was still at disposal (Example 12.3)—this will generate a loss as it is not possible to carry out production during the time used for the investment (The size of the loss can also be illustrated graphically. The reader is encouraged to try this out!). With the fixed asset at disposal, it is possible to gain an immediate profit in connection with the price increase.

The disadvantages of fixed factors are more important than the advantages—everything else being equal. Hence, if it is possible to choose it would normally be an advantage to choose production systems where the possibility of fixation is limited. The possibilities of this are discussed in the next section.

12.5 Limiting the Fixation of the Factors

As described, the fixed production factors are generated when there is a difference between the marginal factor costs of a purchase and a sale. The individual company has no influence on the marginal factor costs of a purchase. The possibility of influencing the fixation of the production factors consists, therefore, primarily in increasing the marginal factor cost of a sale, i.e. acquiring production factors with a high alternative value, which is the same as production factors with a high degree of flexibility.

Assets with very *specific usage* often have a low alternative value. Stable facilities that are built and designed for only one type of livestock can often not be used for other purposes, and the alternative value is therefore low. Such stable facilities will therefore often be fixed factors. The same is true for specialised machines, such as e.g. potato harvesters. However, such machines may be sold and are not physically fixed, such as e.g. stable facilities. However, the problem with such specialised machines is that if the product price decreases, and the producer therefore wishes to reduce the capacity through sale, then there will be many other producers who are in the same situation, and the supply of used specialised machines will be so huge that the price will thus be very low. The alternative cost of keeping the machine is therefore similarly low.

An efficient way of dealing with the fixation of production factors is by *leasing instead of owning*. Depending on the length of the lease period and the contractual conditions, leasing will—everything else being equal—normally provide more flexibility as capacity can be reduced simply be discontinuing the lease. However, the flexibility naturally depends entirely on the contract, and leasing will often be more expensive than owning. Hence, in most cases, there will be pros and cons to be considered.

12.6 The Duration of the Fixation

How long is a production factor a fixed factor?

It is not easy to give a precise and unequivocal answer to this question. In the short run, all factors are fixed as, if the time frame is sufficiently short, it is not physically possible to buy or sell, which in reality means that the purchase price, is infinitely high and the sales price is zero.

However, the longer the time frame, the better the possibilities for achieving a good price when selling. Even relatively fixed factors such as stable facilities and slurry tanks can be *let* to other producers in the long run. If this is the case, then the lost rental income would constitute the marginal factor costs at realisation.

The degree of fixation and the time during which the production factors are fixed, thus, depend entirely on the actual circumstances. The better the possibilities for alternative usages, the smaller the probability that the factors will become fixed, impeding the adjustment in case of changes in price ratios and new technological possibilities.

12.7 Measuring the Fixation of Production Factors

The fixation of production factors can be measured by calculating the company's total value with or without the production factor in question. If the value of the company decreases more at realisation of the production factor than through the revenue generated by the realisation, then the production factor has a certain degree of fixation. The larger the difference between the two amounts, the higher the degree of fixation.[3]

The theoretical argument for using this method to determine the fixation of production factors presupposes knowledge of investment theory, which is beyond the scope of this book.

[3] Assume that the value of the property *with* the asset is MU A, the value *without* the asset is MU B, and the sales price of the asset is MU C. The fixed-price index (FI) can then be calculated as: FI = $(A - B - C)/(A - B)$. If e.g. $C = 0$ then FI is equal to 1, corresponding to an entirely fixed asset. If $C = A - B$ then FI is equal to 0, corresponding to an entirely variable asset.

Chapter 13
Decreasing Sales Curve

13.1 Introduction

The discussion so far has been based on the assumption of perfect competition in the product market—where the producer is a price taker, i.e. can sell unlimited amounts at one and the same price.

While this precondition is realistic in connection with small companies producing standard goods within an industry such as farming, the precondition becomes problematic in connection with larger producers, or producers producing differentiated products/special products for a smaller segment of the market. This could in fact also be problematic for producers who, for some reason or other, have achieved a position whereby they are the only producers of a specific product and have therefore attained a monopoly position in the market. For such producers, the assumption that it is possible to sell any amounts at the same price is of course invalid. The company's sales curve is, under such conditions, equal to the market demand curve (D), which is normally a downward sloping curve as illustrated in Fig. 13.1.

In this chapter, the producer's optimisation of the production (and sales) should be considered under the more general preconditions that the sales price depends on the quantity sold. Perfect competition and monopoly are, in this connection, the two extremes, and these extremes must therefore first be analysed.

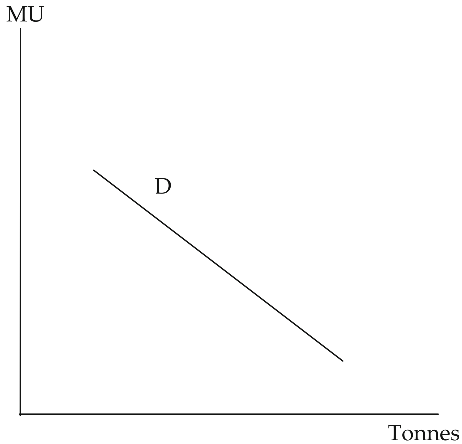
To achieve an understanding of the actual difference between monopoly and perfect competition, including especially the relationship between the company's sales curve and the market demand curve, the conditions of perfect competition will first be examined.

13.2 Perfect Competition as a Reference

The underlying basis for the analysis is the market demand curve for the product that is produced by the company. In the following, this demand curve is assumed to be given by a decreasing curve, as illustrated by D in Fig. 13.1.

S. Rasmussen, *Production Economics*, DOI: 10.1007/978-3-642-14610-7_13,

Fig. 13.1 Market demand curve



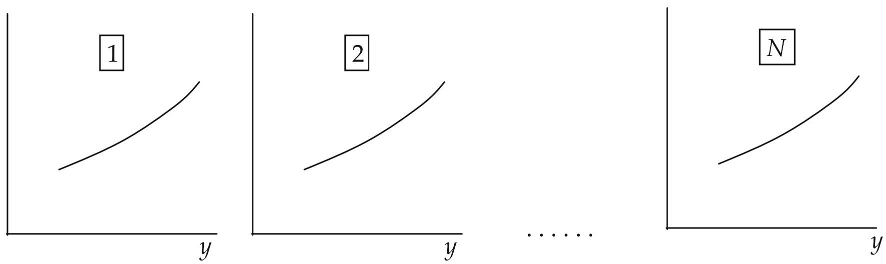


Fig. 13.2 Companies' supply curves (*MC* curves)

Under perfect competition, there are many producers and suppliers of the product *y*. Each of these suppliers has a supply curve corresponding to the company's marginal cost curve. The supply curves for a total of *N* companies are shown in Fig. 13.2.

For the sake of simplification, all companies are assumed to have the same marginal cost curve, and there are assumed to be a total of 1,000 companies ($N = 1{,}000$). In Fig. 13.3 below, all supply curves for these 1,000 companies have been aggregated as all the companies' supply has been added for each price. The result of this is illustrated by the total supply curve (S) for the market, as shown in the coordinate system in the outer right hand side of Fig. 13.3. The same coordinate system also contains the total demand curve D for the market.

As can be seen, there is market equilibrium at a total supply of 1,000 ton, where all 1,000 companies produce the same amount, i.e. 1 ton (t).

What happens if one of the 1,000 producers changes the supply, e.g. by ½ ton?

The left hand side of Fig. 13.4 illustrates company 1 increasing its supply (to s_2) or reducing its supply (to s_1) by ½ ton (±50%) (The other 999 producers continue to supply the original amount without any changes). The implications of this for

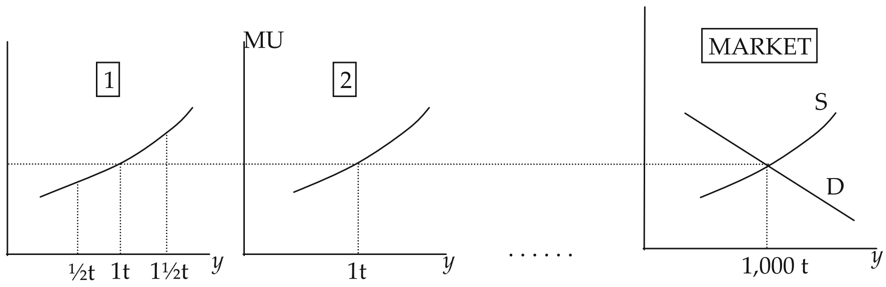


Fig. 13.3 Market equilibrium with many companies

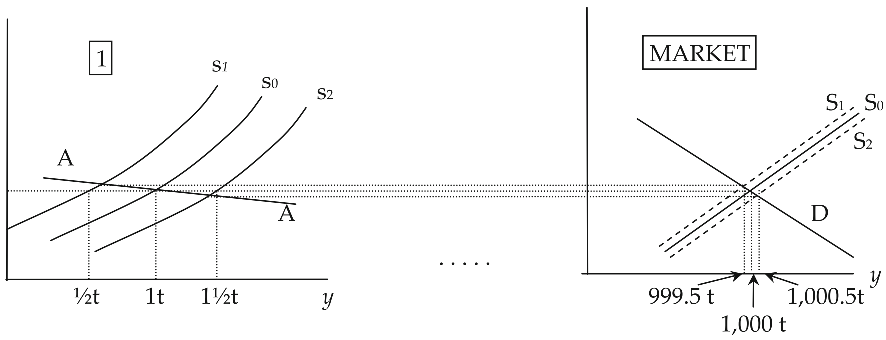


Fig. 13.4 The company's sales curve

the total market supply is illustrated on the right hand side of the figure, which shows that the change in the total market supply by ½ ton (½ pro mille!) only marginally affects the equilibrium price.

The company's *sales curve* is the curve AA. The sales curve is the effective demand function facing the company. If the company produces and sells ½ ton less, the market price is increased marginally. If the company produces and sells ½ ton more, the market price decreases marginally. (For the sake of clarity, the figure on the right is drawn based on an incorrect ratio, and the real price difference between the three situations would—if it was drawn at the right ratio—hardly be detectable). The sales curve is also called *the efficient demand curve facing the firm*. In the following we use the short term *sales curve*.

As can be seen, the company's sales curve is almost horizontal. If the total number of companies in the market moves towards the infinite, the curve AA in Fig. 13.4 moves towards a completely horizontal sales curve.

This is what is understood by *perfect competition in the product market*: The company is facing an (almost) horizontal sales curve, because the amounts produced by the company compared to the entire market are very small.

13.3 Monopoly

We will now look at the other extreme—monopoly.

Under monopoly there is only one producer (supplier) in the market. Consider e.g. company 1 from before. This company has taken over all the other companies in the market and now produces the total market supply, as shown on the left hand side of Fig. 13.5. As before, there are three different examples of the company's marginal cost curves, which are the curves that, under perfect competition, are equal to the company's supply curve and which are here (unchanged compared to Fig. 13.4) referred to as s_0, s_1, and s_2.

However, under monopoly, the company's marginal cost curve is no longer identical to the company's supply curve. A supply of e.g. 500 ton is not based on the condition: product price equal to marginal costs. The right hand side of Fig. 13.5 shows that with a supply of 500 ton it is possible to achieve a price of p_1 in the market in question. At a supply of 1,000 ton, a price of p_0 is achieved. And at a supply of 1,500 ton, a price of p_2 is achieved. And these prices are not directly related to the cost curve of the company.

For companies that constitute a monopoly, there is thus no actual supply curve—a curve showing the relationship between a product price and the quantity that the company is willing to supply. On the other hand, there is an available sales curve—a curve showing the relationship between the quantity supplied and the price achieved. *And this sales curve is, as can be seen in* Fig. 13.5*, equal to the market demand curve*.

Under monopoly, the company's sale curve is, therefore, identical to the market demand curve.

While the sales curve is horizontal under perfect competition, it is decreasing and equal to the market demand curve under monopoly. It is not difficult to imagine that, in between these two extremes, there are various intermediate situations, in which the company's sales curve becomes increasingly horizontal the further the company moves away from the monopoly situation and the closer to perfect competition. One of these intermediate situations is called *monopolistic competition* and this type of market is in fact very common. We will return to this type of market after the treatment of monopoly in the next section.

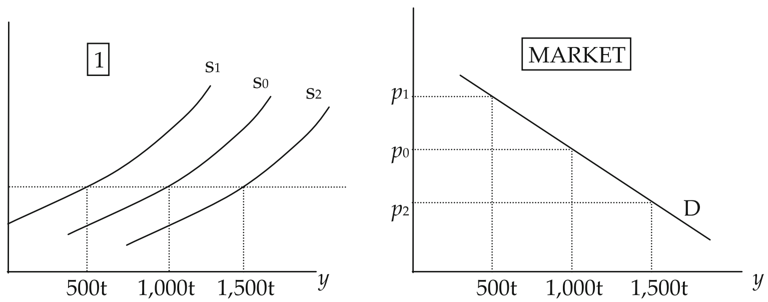


Fig. 13.5 The company under monopoly

13.4 Optimisation Under Monopoly

Under monopoly, the product price is a function of the produced (and supplied) quantity, i.e.:

$$p_y = p_y(y) \tag{13.1}$$

Thus, the company's total profit is:

$$\pi = p_y(y)y - TC \tag{13.2}$$

were TC stands for total costs (see Chap. 5). The profit-maximising production is found by differentiating (13.2) with regard to y and setting the derivative equal to zero. This gives:

$$\frac{\partial \pi}{\partial y} = y\frac{\partial p_y}{\partial y} + p_y(y) - MC \tag{13.3}$$

in which MC are the marginal costs. If the derivative is set equal to zero, the criterion for profit maximisation results in:

$$y\frac{\partial p_y}{\partial y} + p_y(y) \equiv MR = MC \tag{13.4}$$

where MR is the marginal revenue, namely the added value achieved by selling one more unit of y.

The expression $\partial p_y/\partial y$ in (13.4) is equal to the slope of the company's sales curve. Please note that when the slope of the sales curve is zero (perfect competition), then the MR is equal to the product price p_y, and the optimisation criterion is thus equal to the criterion derived in Chap. 9. Hence, perfect competition is a special case of the general model when the company's sales curve is horizontal.

Under monopoly, the company's sales curve is equal to the market demand curve. The expression $\partial p_y/\partial y$ is therefore an expression of the slope of the market demand curve. Normally, the reciprocal expression, i.e. $\partial y/\partial p_y$, is used for the description of the market demand. With regard to the description of the market demand, it is interesting for a company under monopoly to know how much the demand changes when the price increases or decreases. To describe this, the (product) demand elasticity E_d is normally used, as defined by:

$$E_d \equiv \frac{\frac{\partial y}{y}}{\frac{\partial p_y}{p_y}} = \frac{\partial y}{\partial p_y}\frac{p_y}{y} \tag{13.5}$$

If this expression is solved for $\partial y/\partial p_y$ in the following way:

$$\frac{\partial y}{\partial p_y} = \frac{1}{\frac{\partial p_y}{\partial y}} = E_d\frac{y}{p_y} \tag{13.6}$$

and inserted in (13.4), then the marginal revenue (MR) can be expressed as:

$$MR = p_y\left(1 + \frac{1}{E_d}\right) \tag{13.7}$$

With a decreasing demand curve, the demand elasticity E_d is negative. The marginal revenue (MR) is therefore less than the product price p_y. Only when the demand elasticity is very high (the demand curve is (almost) horizontal) is $MR \cong p_y$. The steeper the demand curve, the smaller (the absolute value of) the demand elasticity E_d, and therefore the smaller MR. When the demand elasticity is low (steep slope of the demand curve), the act of producing and selling one product unit more has important implications for the earnings, as not only is the last unit sold at a lower price, but all the preceding units are also affected by the price decrease.

This relationship is illustrated graphically in Fig. 13.6 in which the demand curve is D. Here it is illustrated that when the demand elasticity is high ($|E_d| > 1$), increased production will result in an increase in the total earnings. Consider the production y_1 as the point of reference. If production is increased by dy, then the price will decrease from p_1 by dp, and the total earnings—the rectangle with the area price multiplied by quantity = total earnings—will increase.

When the elasticity is -1, an increase in production will be precisely cancelled out by a corresponding price decrease, so that the total earnings will remain unchanged. Consider the production y_2 as the point of reference. If production is increased by dy, then the price will decrease from p_2 by dp, and the total earnings—the rectangle with the area price multiplied by quantity = total earnings—will remain unchanged. Finally, when the demand elasticity is low ($|E_d| < 1$), then an increase in production will result in a decrease in the total earnings. Consider the production y_3 as the point of reference. If production is increased by dy, then

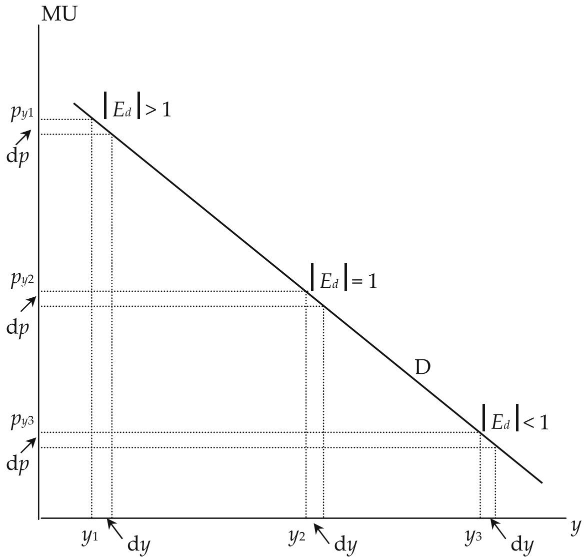


Fig. 13.6 Demand elasticities

the price will decrease from p_3 by dp, and the total earnings—the rectangle with the area price multiplied by quantity = total earnings—will decrease.

In the case of monopoly, the issue might as well be considered from the price side. Whether the producer chooses a quantity and then finds out what price he/she achieves, or whether he/she chooses a price and then finds out what quantity he/she can sell, generates the same result in principle. Therefore, the above analysis might just as well be based on the examination of the implication of price changes.

The criterion for optimal production under monopoly is, thus, that production should be increased as long as the marginal revenue (MR, given by (13.7) is greater than the marginal costs MC. This criterion is not only valid under monopoly but is basically a general criterion for the optimisation of production—regardless of the market type.

Example 13.1 Presume that the demand function is given by the following linear relationship between the quantity y and the price p_y:

$$p_y = a - by \tag{13.8}$$

in which the parameters a and b are positive. The total production value (TR) is given by:

$$TR = p_y y = (a - by)y = ay - by^2 \tag{13.9}$$

and the marginal revenue MR thus by:

$$MR = dTR/dy = a - 2by \tag{13.10}$$

The demand elasticity E_d is given by:

$$\frac{1}{E_d} = \frac{1}{\frac{\partial y/y}{\partial p_y/p_y}} = \frac{\frac{\partial p_y}{p_y}}{\frac{\partial y}{y}} = \frac{\partial p_y}{\partial y}\frac{y}{p_y} \tag{13.11}$$

If (13.8) is differentiated with regard to y and inserted in (13.11), the following demand elasticity expression is produced by isolating E_d in (13.11):

$$E_d = -\frac{p_y}{by} \tag{13.12}$$

which is negative when b is positive.

As can be seen from the comparison of (13.8) and (13.10), the MR curve is twice as steep as the sales curve which is in this case (under monopoly) identical to the demand curve. This is also illustrated graphically in Fig. 13.7.

In Fig. 13.7, the company's cost curves have also been drawn (marginal costs (MC) and average costs (AC)). Using the optimisation criterion (13.8) shows the optimal production to be y_1 where $MR = MC$. The corresponding market price is p_{y1} as this, according to the demand curve D, is the price that can be achieved by

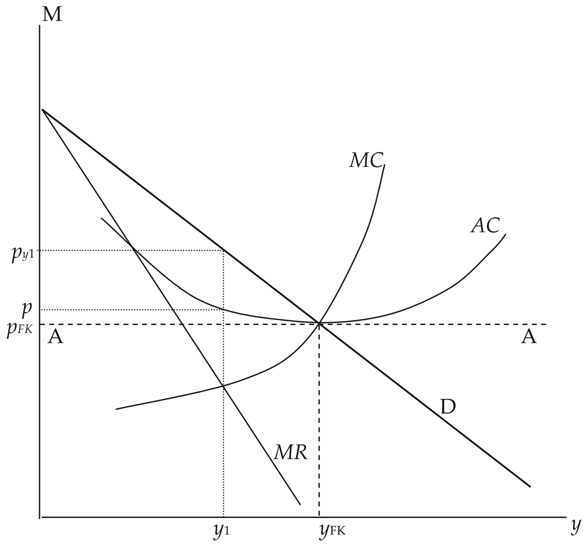


Fig. 13.7 Demand (*D*) and marginal revenue (*MR*)

producing and selling the quantity y_1. Hence, it is optimal for the producer to either choose the sales price p_{y1} (and thus be able to produce and sell the quantity y_1), or to choose to produce and sell the quantity y_1, and thus achieve the price p_{y1}.

Please note that the producer achieves a price which is higher than the production costs. At the quantity y_1, the average total production costs (AC) are equal to p, while the sales price per unit is p_{y1}. This profit is called the *monopoly profit*.

We can now make a comparison with a situation under perfect competition. In Fig. 13.7, the demand curve (D) (the sales curve under monopoly) is in fact drawn so that it goes through the intersection point for the MC and AC curves. Hence, a situation with perfect competition can be directly illustrated as the horizontal sales curve under perfect competition can be achieved by simply pivoting the curve D counter clockwise around this intersection point. This produces the sales curve AA which in fact represents the long run equilibrium under perfect competition, where the product price is equal to the lowest long run costs. As can be seen, the company will, under perfect competition, produce more (y_{FK}) and achieve a lower price p_{FK}. And the profit—the difference between price and average costs—will be zero.

Under market conditions of monopoly, companies will thus—everything else being equal—produce less than under perfect competition, and achieve a higher price.

13.5 Optimisation from the Input Side

While the previous sections of this chapter have been dealing with the company's optimisation as seen from the output side, this section will briefly illustrate how the optimal production is determined from the input side (optimisation of the amount of input x), when the output price depends on the produced quantity.

The criterion for profit maximisation, as seen from the input side under the assumption that the output price is independent of the produced quantity, has previously been derived in Chap. 7. The criterion is:

$$\frac{VMP_1}{MFC_1} = \frac{VMP_2}{MFC_2} = \cdots = \frac{VMP_n}{MFC_n} = 1$$

whereby VMP_i is the value of the marginal product of input x_i ($MPP_i p_y$) and MFC_i are the marginal factor costs for input x_i (see (7.28)).

With a decreasing sales curve, the output price p_y depends on the produced quantity y, which in turn depends on the quantity of added input $\mathbf{x}$. The total value of the production (TVP) is therefore equal to:

$$TVP = p(y(\mathbf{x})) \times y(\mathbf{x}) \tag{13.13}$$

in which $p(y(\mathbf{x}))$ thus is interpreted as the output price p being a function of the production y being a function of the amount of input $\mathbf{x}$.

Differentiating (13.13) with regard to the i'th input x_i produces the marginal product value (MVP_i) which is thus given by:

$$MVP_i = \frac{\partial TVP}{\partial x_i} \tag{13.14}$$

This is an expression of the increased value achieved by the addition of a unit more of input x_i. Calculating the expression by differentiation produces:

$$MVP_i = VMP_i + VMP_i/E_A = VMP_i\left(1 + \frac{1}{E_A}\right) \tag{13.15}$$

Please note that the *marginal product value* (MVP_i) is equal to the well-known expression VMP_i (*the value of the marginal product*) multiplied by a factor (the parenthesis in (13.15)), the value of which depends on the elasticity of sale E_A, which is relative change of sale divided by relative price change. As the elasticity of sale is negative, the marginal product value is thus less than the value of the marginal product ($MVP_i \leq VMP_i$). An extra unit of input thus results in lower value added when the sales curve is decreasing than when it is horizontal. If the elasticity of sale moves towards the infinite (or rather *minus* infinite), the expression in the parenthesis in (13.15) moves towards 1. In this situation, $MVP_i = VMP_i$, perfect competition (the producer is a *price taker* in the product market) is thus found to be a borderline case of the more general market type, i.e. the borderline case where the sales curve is horizontal (see Sect. 13.2).

For the optimal supply of input, the marginal product value should be equal to the marginal factor costs, or the ratio between the two should be equal to 1, i.e.:

$$\frac{MVP_i}{MFC_i} = 1 \tag{13.16}$$

As this is true for all (variable) inputs, it produces the general optimisation criterion:

$$\frac{MVP_1}{MFC_1} = \frac{MVP_2}{MFC_2} = \cdots = \frac{MVP_n}{MFC_n} \tag{13.17}$$

where MFC_i is given in (7.28) and MVP_i is given by (13.15).

Example 13.2 Presuppose a production function $y = f(x_1,x_2) = 6x_1^{0,3}x_2^{0,5}$(as in Example 4.1). The marginal factor costs are presumed to be constant, so that $MFC_1 = w_1 = 8$, and $MFC_2 = w_2 = 12$. It is furthermore presumed that the output price $p = 10 - 0.15y$.

The elasticity of sale is calculated first. The elasticity of sale (see (13.23)) is:

$$E_A = -\frac{1}{0,15}\frac{(10 - 0,15y)}{y} \tag{13.18}$$

Then the marginal product value MVP_i for $i = 1$ and 2 is calculated. Using (13.15) gives:

$$MVP_1 = (10 - 0,15y)(1,8x_1^{-0.7}x_2^{0.5})\left(1 + \frac{1}{E_A}\right) \tag{13.19}$$

$$MVP_2 = (10 - 0,15y)(3x_1^{0.3}x_2^{-0.5})\left(1 + \frac{1}{E_A}\right) \tag{13.20}$$

where by E_A is given in (13.18). If the expression of the production function y is inserted in (13.18), (13.19), and (13.20), and (13.16) is used, the following two conditions for optimal production are generated:

$$\frac{(10 - 0,15 \times 6x_1^{0.3}x_2^{0.7})(1,8x_1^{-0.7}x_2^{0.5})\left(1 + \frac{1}{E_A}\right)}{8} = 1 \tag{13.21}$$

$$\frac{(10 - 0,15 \times 6x_1^{0.3}x_2^{0.5})(3x_1^{0.3}x_2^{-0.5})\left(1 + \frac{1}{E_A}\right)}{12} = 1 \tag{13.22}$$

These two equations with two variables can, in principle, be solved, and the optimal input amount and the production can thus be calculated (not done here).

Dividing (13.19) by (13.20) and using condition (13.17) with $MFC_1 = 8$ and $MFC_2 = 12$, produces the following condition for the optimal combination of x_1 and x_2: $\frac{8}{12} = \frac{1,8}{3}\frac{x_2}{x_1}$ which corresponds to the straight line $x_2 = \frac{24}{21.6}x_1$. Hence, the two inputs should (as was the case in Example 4.1) be used in a constant ratio (linear expansion path). However, we do not know have much input should be applied, but this can be derived by solving (13.21) and (13.22) simultaneously.

13.6 Product Differentiation and Monopolistic Competition

In a market with a large number of firms producing identical products, the sales curve (the demand curve facing the individual producers) is essentially flat, as shown in Sect. 13.2 (perfect competition). The firm must sell its products at the prevailing market price. If the firm raises its price above the market price it loses its entire sale because others are ready to take over at the prevailing price.

If, on the other hand, a firm is the one and only producer of a specific product, it may be able to raise its price without losing its entire sale. Some customers may stop buying the product, buy less, or they may switch to other products, which may to some extent substitute the product in question. But some costumers will still buy the product because they can still afford it or because there are no obvious alternatives. The behaviour of the customers determines the elasticity of the demand curve facing the producer, and the optimal pricing and production is determined as shown in the monopoly case in Sect. 13.4.

If a firm makes a profit producing and selling a specific product, it may be attractive for other firms to enter the industry. The entering firm may not be able, or may not be allowed (by law or regulation), to produce the exact same product. However, they may be able to produce *similar products*—similar in the sense that customers are inclined to switch to this product if the producer of the other (first) product increases the price (too much).

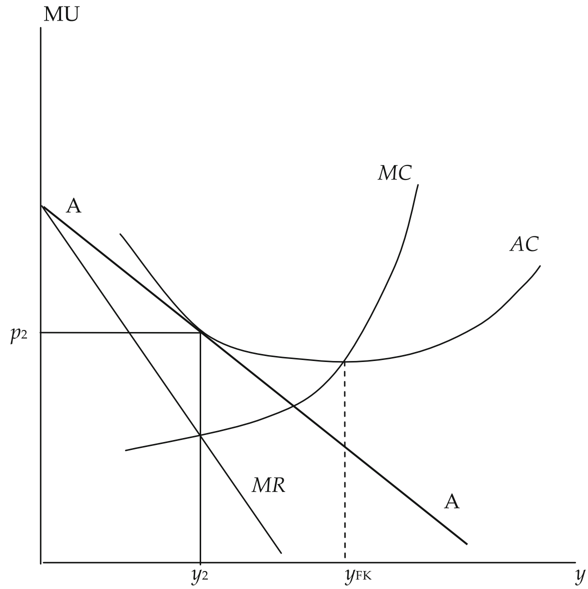
If a firm produces a product, which is difficult for consumers to substitute with other products, then the sales curve will be relatively steep (low demand elasticity). The easier it is to substitute the product, the flatter the sales curve. As the objective of the firm is to increase earnings by increasing the price without losing (too many) sales, firms are motivated to produce similar *but* distinctive products. This is called *product differentiation*, which is what firms do to differentiate their products from those of other firms in the industry.

A market structure such as this, i.e. a market, where it is possible for firms to enter and leave the industry freely, and where products are differentiated, is called *monopolistic competition*. It is monopolistic in the sense that each firm faces a declining sales curve, and it is competitive in the sense that there is fierce competition with other firms, because there is free entry into the industry.

There are many examples of this type of market. The market for furniture is a typical example. In fact, most markets are probably markets with monopolistic competition, with the extreme cases of perfect competition at the one end, and monopoly at the other in their pure forms are in fact rare in practice. However, the limiting cases are nice in the sense that they are easy to analyse. Monopolistic competition is much more difficult to analyse because the results depend on the specific assumptions concerning products, technology and the strategic choices of the firm, etc. However, most microeconomic textbooks present the general result, that the long run equilibrium of markets with monopolistic competition is characterised by the participating firms making zero profit (As in the long run equilibrium case of perfect competition). The model is shown in Fig. 13.8.

The benchmark is an industry in which the individual firms make profit as illustrated in the monopoly case in Fig. 13.7. Due to the prospect of making a profit, other firms enter the industry producing similar but distinctive products. This makes the sales curve (AA) for the firms already in the industry shift inwards. The process of new firms entering the industry and the sales curves shifting inwards continues

Fig. 13.8 Monopolistic competition



until the profit potential is fully exploited, and none of the firms make any positive profit because their sales curves have moved inwards until the point where the profit is zero (price equals average cost), as shown in Fig. 13.8. The long run equilibrium is where the firm produces y_2 units of output and sells the product at a price of p_2 where marginal revenue (MR) is equal to marginal cost (MC) and where the profit is zero.

The criterion for optimal production under monopolistic competition is, in principle, the same as under monopoly, as the demand elasticity (E_d) in the optimisation condition (13.7) is simply replaced by *the elasticity of sale* (E_A) where the elasticity of sale is defined as:

$$E_A = \frac{\frac{\partial y_A}{y_A}}{\frac{\partial p_y}{p_y}} = \frac{\partial y_A}{\partial p_y}\frac{p_y}{y_A} \tag{13.23}$$

and where $\partial y_A / \partial p_y$ is the slope of the company's *sales curve.*

As previously mentioned, an industry with monopolistic competition is one where new firms are free to enter. But what tempts a new firm to enter into an industry in which the long term prospect is to earn zero profit? (The same question could be asked to firms entering industries under perfect competition, where the long term prospects are also zero profit).

An obvious explanation is that the firms entering the industry, or expanding their production within an industry, expect to do better than the (average) firms already in the industry: Firms entering a market under perfect competition expect that they will be able to produce at *lower costs* than the other firms in the industry, and firms entering an industry with monopolistic competition expect that their differentiated products will be able to receive a *higher price* than the other products on the market. Based on these expectations, they enter the industry expecting to make a positive profit.

This approach (strategy) may work in the short run. A new, differentiated product may have special consumer interest for a period after its introduction and it may be possible, therefore, to earn a positive profit in the short run. The same is the case if the firm is able to produce at lower costs than other firms. However, in the long run, other firms will compete by introducing even more advanced (differentiated) products, which the costumers will now turn to, and the sales curve will move inwards. Concerning the low cost strategy, other firms will soon learn how to produce with the same low, or even lower costs thereby eroding the cost advantage.

To continue to generate a positive profit under monopolistic competition it is therefore necessary to continue the process of product differentiation. The ideal case is that the firm is always in front of the others, so that as soon as the competing firms have developed new (differentiated) products, then the firm in question is ready to launch an even more attractive product. A similar approach is necessary if cost minimisation is the competitive strategy of the firm. In this case, the firm should always be ready to look for new cost reducing production methods, which makes the foundation for generating positive profit while others have zero profit.

It is not possible to get customers to pay an excessive price for a product unless they consider it to be *something special*. Therefore, product differentiation is when *a producer is able to add characteristics, which are of value to customers, to a product, which other producers are unable to add to their products*.

The identification of sources for differentiation is an important task for a company manager. In relation to production, i.e. acquisition of inputs, the manufacturing process and the delivery process, there are a number of opportunities which will be briefly mentioned in the following.

13.6.1 Acquisition of Input

The firm may be able to acquire input that other firms are unable to acquire. It may be raw materials or intermediate products that are necessary or important in the production process. A factory producing amber ornaments may have exclusive access to geographical areas where high quality amber can be found, and a producer of TV sets may have a special agreement with a producer of electronic equipment to deliver high quality electronic components. In agriculture, land can be of special quality which provides the foundation for the production of special products, or products of a particular quality (wine). The geographical location of a company can entail special advantages with regard to the *acquisition of input*. The location might entail the possibility of using waste products from nearby industrial companies. *Labour* may require a specific education which provides the foundation for the execution of a particular (and difficult) production process. *The climate* might provide possibilities which other companies do not have. It might be possible to make an agreement with a supplier of a particular input so that they won't deliver to other producers (Such agreements cannot, of course, be entered into if they are in breach of the law).

13.6.2 The Production Process

The differentiation can be based on a particular production technology or special facilities. An example is the production of special brands of cheese or ham in France or Spain which are based on the storage of the products in mountain caves. Furniture producers may differentiate their products by making them according to a special design. Producers may use special equipment to monitor the production process, and based on this, be able to provide guarantees for specifications including, for instance, particular quality requirements. This may be part of a process of building up a brand. To be of lasting value, it is important that the differentiated production method cannot be imitated by other firms. Indeed, firms may invest considerable resources in order to protect their unique production process from industrial espionage.

13.6.3 Delivery of Output

The differentiation on the output side may e.g. consist of the correct *timing*, so that the purchaser/consumer receives his/her product precisely when it is needed. It can also be a question of delivery to the right *premises,* and of delivery in units, sizes, or the like, which are well suited for *transportation*. A *guarantee for delivery* of the product in certain amounts, with a *guarantee of specifications,* delivered at the agreed time, can also be of importance to the customer and is therefore potentially useful as part of a product differentiation.

Differentiation is often cost-intensive; however, it does not need to be so. Company resources may already exist that could be used as the basis for differentiation. And if there are resources which other companies do not have access to, then a small monopoly has already been created, and it is possible—if used correctly—to achieve a monopoly profit.

The problem in connection with a product differentiation arises if other companies simply imitate. If this is the case, an expected extra profit can disappear as others are taking over and might even do a better job. It is therefore important to base it on elements that the other companies do not have immediate access to.

This subject is too comprehensive for a thorough discussion here. Please refer to the extensive literature on this subject. A classic within this subject is Porter's book from 1985 (Porter 1985).

Reference

Porter, M. (1985) *Competitive advantage*. New York: The Free Press.

Chapter 14
Production Over Time

14.1 Introduction

Time is a key input in all kinds of production. However, the subject of using time for production is more relevant in some industries than in others. The distinguishing feature of agriculture and other biologically based industries compared to other industries is the fact that production often takes quite a long time. Just think of the production of wood from beech trees, which first reach maturity after 50–100 years of growth!

How is the time consumption optimised in a production process? What is the decisive factor for determining when to send an animal, e.g. a slaughter pig or a beef calf, to the slaughterhouse? Should it be sent to the slaughter house one week earlier or later? What is the decisive factor for how many years the strawberry plants should be used before removing and replacing them with other plants? How much time should pass before a stand of trees is felled so that new trees can be planted? It is issues like these that will be discussed in this chapter.

First, the issue is considered within the framework of an entirely general model, in which time is considered along the same lines as other inputs. This approach shows that time in production, in principle, can be treated within the framework of an already developed model structure as an extra dimension by simply introducing it when describing inputs, outputs and prices: To all inputs, outputs and prices are added an extra index, time, stating the time periods that are relevant for the input or output in question or the time period that the price in question refers to. This is in line with the general definition of a good, after which two goods are defined as being different goods if they are different with regard to kind, location or time. A bottle of Coke on sale at the Town Hall Square in Copenhagen on the 1st of January is, thus, a different good to a bottle of Coke on sale at the Town Hall Square in Copenhagen on the 30th of June! By using this distinction between goods (input and output), depending on the time at which they are available, all production-related issues involving time can, in principle, be handled.

S. Rasmussen, *Production Economics*, DOI: 10.1007/978-3-642-14610-7_14,

When talking about prices at different points in time, time preferences and, hence, time-related costs in the form of interest must be taken into consideration. The criteria for the optimal production period, which are developed in the form of *replacement calculations* later in the chapter, require a certain level of basic knowledge of interest calculation. In the appendix to this chapter you will find a brief introduction to the formulas used in connection with interest calculation.

14.2 General Theory

As a basis for this chapter, it is examined whether time, as an input, can be treated as other (variable) inputs using one of the well-known optimisation principles.

In Fig. 14.1, production is represented as a function of time. If the traditional optimisation principle is used, the optimal time application is represented as t_{opt}, as the ratio between price of time (w_{time}) and the product price (p_y) is given by the slope of the dotted line.

Figure 14.2 shows that the optimal combination of the variable input fodder (x_2) and the application of time (x_1) for the production of a beef calf of 400 kg is B units of fodder and A units of time.

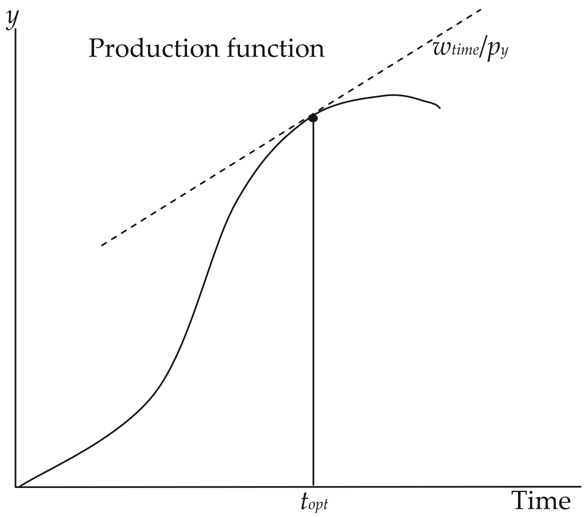
As such, there seem to be no fundamental problems with the treatment of time as a normal variable input. Hence, the production function could be written as:

$$y = f(x_1, x_2,\ time \mid x_3, \ldots, x_n) \tag{14.1}$$

whereby x_1, x_2, and *time* are variable inputs, and x_3, …, x_n are fixed inputs.

The problem is, however, that when time is introduced, then not only does the length of the production period need to be considered, but also the *distribution*

Fig. 14.1 Optimisation with time as input



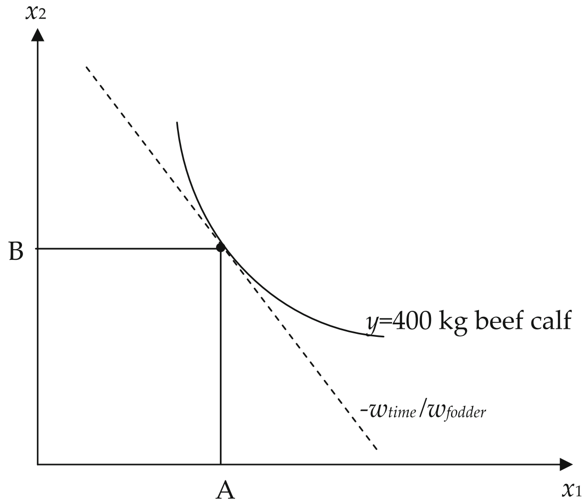


Fig. 14.2 Optimisation of time and feed

over time of the other inputs. It is, after all, not expedient to e.g. give all the fodder to a beef calf on the 25th day of the growth period. The animal would have been dead for a long time by then and the fodder would not do it any good!

Therefore, it has to be decided how the fodder (and the other inputs) should be precisely distributed over time.

In general micro-economic theory, there are no fundamental problems in allocating inputs over time. A given input at the time t_1 is thus another input if it is added at the time t_2. The production function in (14.1) can therefore be written as:

$$y = f\left(x_1^1, \ldots, x_1^T, x_2^1, \ldots, x_2^T \middle| x_3^1, \ldots, x_3^T, \ldots, x_n^1, \ldots, x_n^T\right) \qquad (14.2)$$

where x_j ($j = 1, 2$) are the variable inputs, and x_j ($j = 3,\ldots,n$) are fixed inputs, and where the top signs refer to the time periods for the use of the input in question. The time T refers here to the absolute last time period when the supply of input would be relevant (T is the maximum, technically possible production period).

Please note that the choice of a production period of t year(s) ($t < T$) in the model (14.2) can be formulated as:

$$\begin{aligned} y &= f(t) \\ &= f\left(x_1^1, \ldots, x_1^t, 0, \ldots, 0, x_2^1, \ldots, x_2^t, 0, \ldots, 0 \middle| x_3^1, \ldots, x_3^t, 0, \ldots, 0, \ldots, x_n^1, \ldots, x_n^t, 0, \ldots, 0\right) \end{aligned} \qquad (14.3)$$

as the application of input in the periods *after* the period t is set equal to zero.

Hence, the choice of the length of the production period can formally be described as the choice of the amount of input used in each individual sub-period. The production period ends (formally and actually) when the addition of all inputs are set equal to zero.

For predetermined values of input in each sub-period, the profit from the production can now be calculated as a function of the time *t*. If the profit is referred to as π, it can simply be calculated as:

$$\begin{aligned}\pi = p_y^t f\big(x_1^1, \ldots, x_1^t, 0, \ldots, 0, x_2^1, \ldots, x_2^t, 0, \ldots, 0 \big| x_3^1, \ldots, x_3^t, 0, \ldots, 0, \ldots,\\ x_n^1, \ldots, x_n^t, 0, \ldots, 0\big) - w_1^1 x_1^1 - \cdots - w_1^t x_1^t - w_2^1 x_2^1 - \cdots - w_2^t x_2^t - FC\end{aligned} \tag{14.4}$$

as it is presumed that the output *y* is sold at the price p_y^t at the end of the production period. w_i^k is the price of the input *i* in the period *k* measured in the same monetary unit as p_y^t, i.e. in year *t* MU. Maximisation of the profit for the relevant production process/the relevant asset is found when the derivative of π with regard to *t* equals zero. Differentiating (14.4) with regard to *t* and setting the derivative equal to zero produces the optimal condition (difference calculation, as *t* is discrete here):

$$MR(t) = MC(t) \tag{14.5}$$

in which *MR*(*t*) is the marginal revenue (revenue in one time period) and *MC*(*t*) are the marginal costs (costs in one time period), calculated as:

$$MR(t) = p_y^{t+1} f(t+1) - p_y^t f(t) \tag{14.6}$$

$$MC(t) = w_1^{t+1} x_1^{t+1} + w_2^{t+1} x_2^{t+1} \tag{14.7}$$

respectively, whereby *f*(*t*) is defined in (14.3).

Hence, the optimal production time, when *maximising the total profit for the assets/production process*, is found when the *marginal profit with regard to time is zero,* which can also be expressed as in (14.5), namely in the way that the marginal revenue with respect to time should be equal to the marginal costs with respect to time.

The optimisation illustrated here presupposes that the aim is to maximise the profit for the relevant production process/the relevant asset. However, this is not necessarily a correct formulation if the decision maker's aim is in fact to *maximise the profit per time unit*.

The profit per time unit is given by $\pi(t)/t$. If *t* is to be expressed in days, the profit per year equals $(\pi(t)/t) \times 365$. If the aim is to maximise the profit per year, the optimisation would consist of the following task:

$$\mathrm{Max}(\pi(t)/t) \times 365$$

which is the same as:

$$\mathrm{Max}\ (\pi(t)/t) \tag{14.8}$$

as the number 365 is simply a constant.

Maximisation of the average profit as described in (14.8) gives another optimal (shorter) production period than the maximisation of the total profit. This is examined in more detail in the following.

14.3 The Replacement Calculation

14.3.1 Introduction

When addressing the issue of the lifetime of assets in general, it is important to distinguish between the *technical/biological lifetime* and the *economic lifetime*.

The *technical lifetime* of a machine can be very long. This of course depends on the maintenance; however, there are examples of machines getting very old before they had to be scrapped because they no longer functioned. Assets such as livestock can also reach a very high age before they die of natural causes. The same is true for fruit and berry cultures.

The *economic lifetime* is the time period in which it is profitable to keep the asset from an economic point of view. The economic lifetime is often considerably shorter than the technical/biological lifetime. This is due to the fact that the economic return on the assets we are dealing with here is reduced with increasing age. However, technological developments can also result in new and improved assets entering the market so that the older ones become outdated.

Depending on the type of asset and the conditions, a number of different reasons for the replacement can be outlined:

- Decreasing returns
- Increasing costs
- Changed price ratios
- Decreasing efficiency
- Better assets on the market
- Insufficient capacity
- A combination of reasons.

With the beef calf as an example, it is well-known that maintenance fodder increases with the increase in weight (increasing costs). The marginal gain will also decrease when the animal passes a certain age (decreasing marginal returns). In connection with machines, it is well-known that older machines are more expensive regarding maintenance and repair than newer ones.

The reasons for the replacement of assets can, in principle, be divided into two types, depending on the effect in question. In this connection, the replacement is said to take place due to the *age effect* or the *calendar effect*. It can of course also be due to a combination of these two.

The age effect refers to the situation in which the incoming and outgoing payments for an asset depend solely on the age of the asset. Hence, the price of a new machine is assumed to be the same whether it is bought in the year 2005, or the year 2009. The same is true for the scrap value, the size of which depends on the age of the asset and not the calendar time. Also, the current operating costs are presumed to be unambiguously determined by the age of the asset.

Regarding machines and other technical assets, the age effect is also what we refer to as *wear and tear*. Regarding biological assets, it is also possible to refer to "wear and tear," however, to be more precise, this is usually referred to as biological/physiological aging.

The calendar effect refers to the situation in which the incoming and outgoing payments for an asset depend solely on the calendar time of the payments. Hence, the price for a beef calf depends solely on whether it has been acquired in 2005, or in 2010. The age does not matter. Similarly, the amount of the fodder costs depends on the calendar year (calendar time) in which they occur, and not the age of the animal.

In practice, we often see a combination of both age and calendar effect (wear and tear and obsolescence). However, the age effect is typically the dominating element which is why the analysis of the models presented below primarily *presupposes the age effect.*

There will, on the other hand, be no examples of models that presuppose both the age and the calendar effect. The reason for this is that no operational replacement criteria can be derived analytically under such preconditions. The solution to such problems requires the use of algorithmic methods, such as e.g. Dynamic Programming, a subject which is not covered in this book.

In the following, the replacement problems in connection with biological assets are discussed. Here, the problems will be treated based on the assumption that the producer wishes to carry out the replacement in a way that maximises the profit for these assets. The replacement of machines and similar technical assets will not be discussed as this belongs to the subject of investment planning.

14.3.2 Optimal Termination Time for One Individual Asset

The discussion in this chapter is related to assets where the problem is related to the objective of maximising the profit of an asset. Within agriculture, this is typically assets such as livestock or plant cultures such as e.g. fruit and berries.

The net present value for a one-time investment (see Appendix) can be written as:

$$K_0(n) = -A + \sum_{t=1}^{n} b_t(1+i)^{-t} + S_n(1+i)^{-n} \tag{14.9}$$

in which $K_0(n)$ is the net present value for an asset that is used in n time units, A is the asset purchase price, b_t is the net operational payment for the year t, i.e. the difference between incoming operational payment in period t (v_t) and outgoing operational payment in the year t (u_t) ($b_t = v_t - u_t$) (b_t is paid at the *end* of each period t), i is the interest rate per time unit and S_n is the salvage value (the sales value) for an n year old asset.

According to neoclassical investment theory, the net present value can be used for ranking alternative investments. This is also the case in connection with problems of replacement. The alternative investments are, however, not alternative

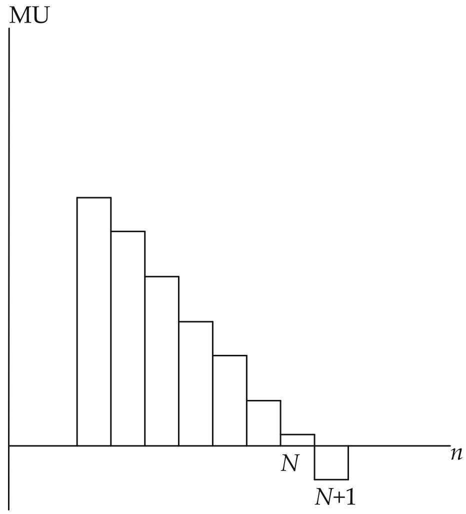


Fig. 14.3 Extra profit as a function of lifetime

assets, but rather alternative lifetimes (n) for a given asset. Hence, the task will consist of determining the lifetime n that maximises the net present value of the asset.

The desired optimal lifetime is referred to as N. For N to be optimal, i.e. give rise to the highest net present value, the following must be true:

$$K_0(N) - K_0(N+1) > 0 \tag{14.10}$$

$$K_0(N) - K_0(N-1) > 0 \tag{14.11}$$

The conditions can be outlined graphically in Fig. 14.3.

These two formulas show the conditions for optimality.[1] If the expression for $K_0(n)$ in (14.9) is inserted in the formula (14.10) and (14.11), the following conditions are derived:

$$b_{N+1} - iS_N - (S_N - S_{N+1}) < 0 \tag{14.12}$$

$$b_N - iS_{N-1} - (S_{N-1} - S_N) > 0 \tag{14.13}$$

The left hand side of (14.12) shows the extra profit generated by keeping the asset in one more period from t_N to t_{N+1}. This extra profit consists of the net payment b_{N+1} with a deduction of the interest on the salvage value S_N in one period, and with a deduction of the decrease in the salvage value of the asset resulting from keeping the asset in this period.

The left hand side of (14.13) similarly shows the extra profit generated by keeping the asset in the period from t_{N-1} to t_N.

[1] These conditions explicitly express that we are looking for a maximum. Here, and in the following, it is assumed that there is only one maximum (local maximum = global maximum).

As can be seen in (14.12) and (14.13), the optimal termination time is the time where the expected extra profit generated by keeping the asset for one more period becomes negative. Similarly, it can be concluded that the asset should be kept as long as the extra profit of keeping it for one more period is positive.

The criterion can be outlined as shown in Fig. 14.3.

The criterion shown here can also be derived by the use of differential calculus for optimisation. This requires, however, that the payments are presented as a continuous flow, and not as discrete amounts as in (14.9). The interest rate is also required to be continuous.

If the interest rate is $i100$ percent per period, the corresponding *interest intensity* δ by a continuous accrual of interest is equal to:

$$\delta = \ln(1 + i)$$

By a continuous accrual of interest, the expression $(1 + i)$ thus changes to e^{δ}. Similarly:

$$(1 + i)^n = e^{\delta n}$$

$$(1 + i)^{-1} = e^{-\delta}$$

$$(1 + \mathrm{i})^{-\mathrm{n}} = e^{-\delta n}$$

Under the precondition of continuous time, (14.9) can therefore be written as:

$$K_0(n) = -A + \int_0^n b(t)e^{-\delta t}dt + S(n)e^{-\delta n} \tag{14.14}$$

This expression is maximised with regard to n by differentiating with regard to n and setting the derivative equal to 0. If the derivative of $K_0(n)$ with regard to n is referred to as $K_0{}'(n)$, and the derivative of $S(n)$ with regard to n is referred to as $S'(n)$, this gives:

$$K_0'(n) = b(n)e^{-\delta n} + S'(n)e^{-\delta n} - \delta S(n)e^{-\delta n} \tag{14.15}$$

which for optimal $n = N$ is equal to zero. If the expression in (14.15) is set equal to zero and $e^{-\delta n}$ is eliminated, the condition can be written as:

$$b(N) - \delta S(N) + S'(N) = 0 \tag{14.16}$$

This is an expression of the marginal profit with regard to time. The expression consists of the net payment $b(N)$ with a deduction of the salvage value interest and with the addition of the change in the salvage value. (Please note that $S'(N)$ can be positive or negative. For livestock it will normally be positive, as this represents growth.)

Hence, for an optimal $n = N$, the marginal profit with regard to time should be 0. As this concerns the maximisation of the net present value, the asset should thus

Table 14.1 Calculation of the optimisation of the production period for slaughter animals

Animal age (days)	Animal weight (kg)	Sales price, MU (S_n)	Period costs for fodder etc. MU (u_n)	Extra profit for period, MU ($-u_n - i_n S_n - (S_n - S_{n+1})$)
215	250	1813	214	MU −214 −25 + 362 = 123
257	300	2175	249	MU −249 −31 + 363 = 83
300	350	2538	280	MU −280 −39 + 342 = 23
346	400	2880	320	MU −320 −44 + 360 = −4
392	450	3240	–	–

Note: The interest rate per year is estimated at 12%. The interest rate from a given period i_n is therefore calculated as $i_n = 0.12$ (number of days/360)

be kept as long as the marginal profit is positive, and be sold when the profit becomes 0 or negative. The criterion is of course completely similar to the discrete situation mentioned earlier and can therefore be outlined, as shown in Fig. 14.3, by simply replacing the staircase-shaped curve with a smooth curve representing the left hand side (the marginal profit) as a function of time. The optimum is found where this curve intersects the horizontal axis.

The criterion is clearly logical: *An asset should be kept as long as another time period contributes to positive earnings (marginal earnings). It is equally obvious that the asset should be disposed of as soon as the marginal earnings become negative*.

Example 14.1 The criterion will now be illustrated with an example: The aim is to determine when an animal should be sent to the abattoir to be slaughtered.

The data and calculation for the example are outlined in Table 14.1.

As can be seen from the calculations, the extra profit is positive all the way up to the age of 346 days, when the animal weighs 400 kg. The extra profit generated by keeping the animal for another period (up to 450 kg) is however negative (MU −4). The animal should therefore be send to the abattoir when it weighs 400 kg after 346 days.

14.3.3 Optimal Replacement Time for Replacing an Asset with a New One

The problem discussed in the previous section presupposes that the asset in question is not to be replaced by a new one. The temporal planning of the asset can therefore be performed without considering any subsequent assets, as the initiation of any new assets is not affected by the termination time of the previous asset. This relationship can be illustrated as shown in Fig. 14.4.

Here, the time T symbolises the earliest possible initiation time for a possible new asset.

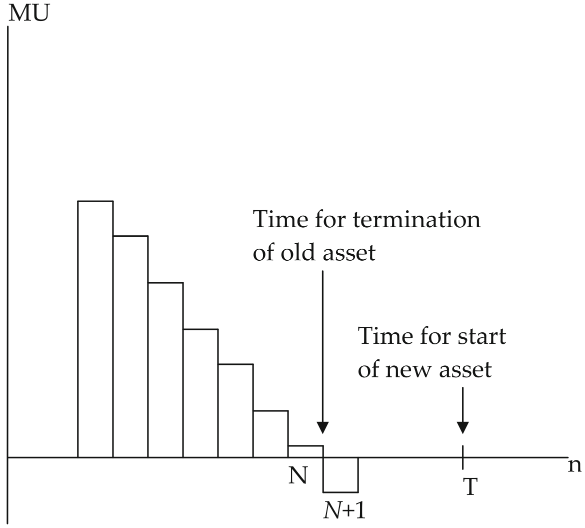


Fig. 14.4 Extra profit as a function of lifetime

A couple of examples are included to illustrate situations in which this is the case:

- A farmer who produces slaughter calves has idle stable capacity. When a calf is sent to the abattoir, the box will under all circumstances be empty for a shorter or longer period of time before it is occupied by a new calf.
- A greenhouse gardener who grows tomatoes terminates production for the year at a time during the autumn. Production of a new batch of tomatoes is not started until February and the greenhouse is empty until then.

In both examples, optimisation of the termination time for the asset should be carried out according to the criterion derived in the previous section, i.e. the asset (the calf, the tomato culture) should be kept as long as the extra profit per time unit is positive.

However, the problem changes nature when the new asset is to immediately replace the old asset, or—expressed more generally—when the initiation time for a new asset depends on the termination of the old asset. If this is the case, the calculation must include the precondition that a continuation of the old asset entails a postponement of the new asset and, thus, a postponement of the corresponding earnings. This results in a loss of interest which should be included in the determination of the replacement time.

The problem can be formulated as:

$$\max_n \{KK_0(n)\}$$

where

$$KK_0(n) = K_0(n) + T(1+i)^{-n} \tag{14.17}$$

The net present value of the first asset is referred to as $K_0(n)$ as previously. This asset must at some point n be replaced by a new asset. The new asset may then, in turn, be replaced by another new asset etc. Which new asset is chosen, and how many there might be, is not of interest right now. Rather, it is simply established that the present (first) asset will be followed by one or more assets.

The net present value of this line of *new* assets calculated for the time t_n is referred to as the *terminal value* and is represented by T in the formula (14.17). The net present value of the entire line (the present (first) asset plus the line of new assets) calculated for the time t_0 is therefore equal to $KK_0(n)$ in the formula (14.17).

The maximum of (14.17) with regard to n entails that the following should be true for an optimal $n = N$:

$$KK_0(N) - KK_0(N+1) > 0 \tag{14.18}$$

$$KK_0(N) - KK_0(N-1) > 0 \tag{14.19}$$

If the right hand side of (14.17) is inserted in the formula (14.18) and (14.19), and if the expression for $K_0(n)$ given in the formula (14.9) is used, it results in the following *conditions of optimality:*

$$b_{N+1} - iS_N - (S_N - S_{N+1}) < iT \tag{14.20}$$

$$b_N - iS_{N-1} - (S_{N-1} - S_N) > iT \tag{14.21}$$

As can be seen, the left hand sides in (14.20) and (14.21) are the same as the conditions shown in (14.12) and (14.13), respectively, in Sect. 14.3.2. In (14.20), it is the extra profit when keeping the asset in the period from t_N to t_{N+1}, and in (14.21) the extra profit when keeping the asset from t_{N-1} to t_N.

The conditions (14.20) and (14.21) entail that it is optimal to carry out the replacement when the expected extra profit for the coming period is less than the interest i of the terminal value T. The asset should, on the other hand, be kept as long as the extra profit is greater than the interest of the terminal value.

The criterion can be outlined as shown in Fig. 14.5. As can be seen, the requirement to replace the old asset with a new one entails that replacement should take place earlier than if there is no direct requirement to replace the asset with a new one. In the example in the figure, the replacement should take place 3 periods earlier. However, as can be seen, this depends on both the size of the interest, as well as the terminal value. A higher terminal value will—everything else being equal—imply that the asset should be replaced earlier.

The optimality criterion (14.20) and (14.21) can, in the same way as mentioned in Sect. 14.3.2, be derived using a continuous time model. Under the precondition of continuous accrual of interest, the criterion shown can be derived by use of differential calculus as in (14.14–14.16) to produce the following optimisation condition as a result:

$$K'_N(N) = \delta T \tag{14.22}$$

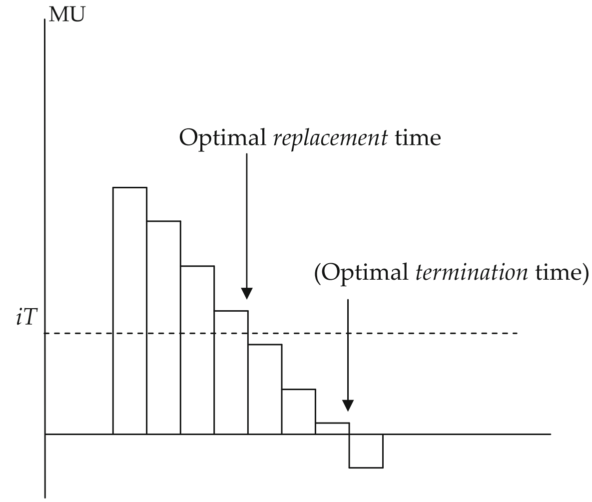


Fig. 14.5 Extra profit as a function of lifetime

in which $K'_N(N)$is the marginal profit of the first asset calculated for the time t_N. This value corresponds to the expression in the formula (14.16). The criterion is of course the same as shown in discrete form in (14.20) and (14.21) and shows that the marginal profit should be equal to the interest of the terminal value for an optimal $n = N$.

Practical usage of the criterion derived here presupposes—apart from the calculation of the marginal profit—calculation of the terminal value T. This may cause problems; however, an operational expression can be calculated based on simplified assumptions.

The problem is that the calculation of the terminal value presupposes that the net present value of all the subsequent assets can be calculated and that the dates for the introduction of each of these subsequent assets are known.

However, the net present value for these assets cannot be calculated without a prior knowledge of their optimal lifetime. And it is, furthermore, necessary to know the optimal lifetime for each asset to be able to discount by the correct number of periods.

$$T = K_0(n_1) + K_0(n_2)(1+\mathrm{i})^{-n_1} + K_0(n_3)(1+\mathrm{i})^{-(n_1+n_2)} + \cdots \qquad (14.23)$$

The problem is outlined in the above formula. T is the terminal value, i.e. the net present value of all the *new assets* at the time t_n-the date of the termination of the present (first) asset.

The calculation of this expression presupposes the estimation of n_1, n_2, n_3,…., i.e. the optimal lifetimes for the first, second, third etc. of the new assets. Hence, the calculation of the terminal value results in a large scale replacement problem.

The problem can be solved, however, by introducing the following preconditions: *The time horizon is infinitely long*[2] and all new assets, which over time will replace each other, are *identical*.

Under these preconditions it is possible to show that all the new assets will have the *same optimal lifetime*. If this lifetime is referred to as n_1, the terminal value can thus be written as:

$$T(n_1) = K_0(n_1) + K_0(n_1)(1+i)^{-n_1} + K_0(n_1)(1+i)^{-2n_1} + \cdots \tag{14.24}$$

This expression can be reduced to:

$$T(n_1) = \frac{1}{1-(1+i)^{-n_1}} K_0(n_1) \tag{14.25}$$

It is however necessary to find the value of n_1 that maximises T. If the expression (14.25) is written based on continuous time as:

$$T(n_1) = \frac{1}{1-e^{-\delta n_1}} K_0(n_1) \tag{14.26}$$

then this value can be found by use of differential calculus. Differentiating (14.26) with respect to n_1 and setting the derivative equal to zero produces the following condition for an optimal value of $n_1 = N_1$:

$$K_0'(N_1) = \frac{\delta K_0(N_1) e^{-\delta N_1}}{1-e^{-\delta N_1}} \tag{14.27}$$

When multiplying by $e^{\delta N_1}$ which moves the marginal profit $K_0(N_1)$ to the time t_{N_1}, (14.27) can be written as:

$$K_{N_1}'(N_1) = \frac{\delta}{1-e^{-\delta N_1}} K_0(N_1) \tag{14.28}$$

The left hand side in (14.28) represents the marginal profit with regard to time. The right hand side in (14.28) represents the average profit for each individual asset expressed as an annuity. This becomes clear when rewriting (14.28) to a discontinuous formula, which gives:

$$K_{N_1}'(N_1) = \frac{i}{1-(1+i)^{-N_1}} K_0(N_1) \tag{14.29}$$

[2] For each individual decision maker the time frame is, of course, limited and this precondition can therefore seem unrealistic. The precondition should, however, not be taken too literally, but only be seen as a reference to the existence of *many* new plants. In practice, the precondition is "correct" if the decision maker's actual decisions are unaffected by the conditions at the end of the time horizon.

As can be seen, the expression $\frac{i}{1-(1+i)^{-N_1}}$ is the well-known annuity factor $a^{-1}_{\overline{N_1}|i}$ which, when multiplied by the net present value, is used to calculate the average profit per period for an asset (see Appendix).

Hence, the optimal lifetime for the new assets is the lifetime $n_1 = N_1$ which fulfils (14.29), i.e. where the marginal profit is equal to the average profit. This is illustrated in Fig. 14.6.

As can be seen, the marginal profit is equal to the average profit precisely where the average profit assumes its maximum. Hence, the criterion (14.29) is equal to the statement that *the optimal lifetime for the new assets is the lifetime that maximises the average profit of each individual asset*.

It is now possible to calculate the desired terminal value. According to (14.26), it is equal to:

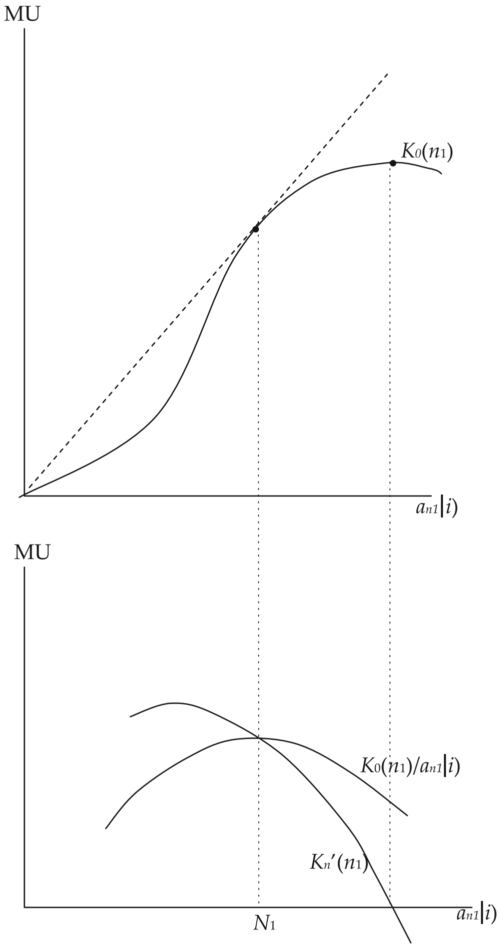


Fig. 14.6 Net present value, average profit, and marginal costs of an asset

$$T = \frac{1}{1-(1+i)^{-N_1}} K_0(N_1)$$

where N_1 is the value of n_1 that maximises the average profit given by:

$$K_0(n_1)\frac{i}{1-(1+i)^{-n_1}}$$

and $K_0(n_1)$ is the net present value of a new asset.

Inserting this expression for the terminal value in the conditions for optimality (14.20) and (14.21) produces:

$$b_{N+1} - iS_N - (S_N - S_{N+1}) < K_0(N_1)\frac{i}{1-(1+i)^{-N_1}} \tag{14.30}$$

$$b_N - iS_{N-1} - (S_{N-1} - S_N) > K_0(N_1)\frac{i}{1-(1+i)^{-N_1}} \tag{14.31}$$

Under the precondition of an infinitely long time horizon and under the precondition of the chain of new assets being identical, *the use of the present asset should continue as long as the extra profit for a period is greater than the (maximum) average profit per period for the new asset, and the asset should be replaced as soon as the expected extra profit is smaller than the average profit of the new asset.* In Fig. 14.7, the criterion has been outlined graphically. Figure 14.7 also shows that the optimal replacement time depends on the performance of the

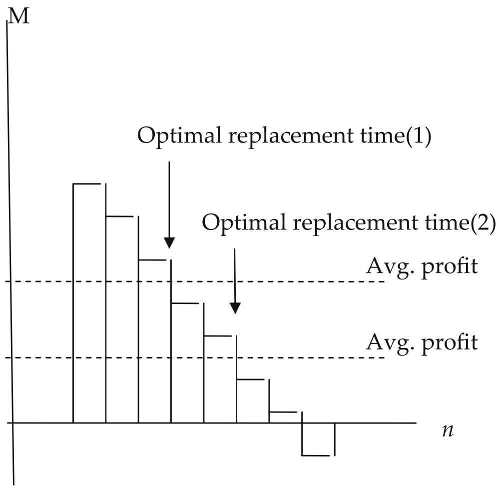


Fig. 14.7 Determination of replacement time

new asset. The higher the average profit for the new asset, the earlier the new asset should be introduced. Hence, a technological development which would mean that the average profit for new assets would increase from (2) to (1) will result in the earlier replacement of the present asset with a new asset.

Example 14.2 Replacement of pigs Table 14.2 illustrates the use of the replacement criterion derived here.

The example illustrates a finisher pig production in which a sty/section/stable is completely emptied of pigs before a new batch enters. The question is: when should the last part of the existing batch be delivered? In the example, calculations have been carried out to determine whether delivery of the new batch should occur now or later. The calculations have been made on a weekly basis under the precondition that delivery can take place once a week.

The calculations show that it is not profitable to deliver the batch in the week in question. The extra profit achieved by keeping the batch for another period is MU 958, while the average profit of the new batch is MU 219 per week.

Please note that it is only possible to decide whether to deliver the batch now, or to wait. It is neither necessary nor expedient to decide whether the batch should be delivered next week, or in 2 or 3 weeks time etc. The conditions may change with regard to expectations, and any new information (such as a change in prices or interest) should be included in the calculations when they are repeated in a week's time to decide whether to deliver the batch or to wait.

Table 14.2 Example, pigs in batch operation

Existing batch:		
Sales value now:	MU 52,000	
Sales value in a week:	MU 56,000	
Marginal revenue:		MU 4000
Fodder	MU 2,800	
Veterinary, utilities etc.	MU 60	
1 week interest of MU 52,000 (0.35%)	MU 182	
Marginal costs:		MU 3,042
Marginal profit		MU 958/week
New batch:		
Purchase of weaner pigs	MU 25,000	
Purchase of fodder	MU 32,000	
Veterinary, utilities etc.	MU 1,500	
Sale, finisher pigs in 3.5 months	MU 65,000	
Discount to net present value (interest = 1.5% per month)	MU 61,700	
Total profit, present value:		MU 3,200
Production time: 15 weeks		
Interest per week: 0.35%		
Avg. profit: $3200 \times 0.0035/(1 - (1 + 0.0035)^{-15})$ = MU 219/week		

Appendix

Interest Calculation

With an interest rate of i, an amount of a MU at the beginning of year 1 (time t_0) is, thus, equivalent to an amount of $a(1 + i)^n$ MU at the end of year n. Similarly, an amount of b MU at the end of year n is equivalent to an amount of $b(1 + i)^{-n}$ at the beginning of year 1.

If, during n years, there is a series of net payments (these can be negative or positive) of $b_0,\ldots, b_{n-1}$ at the beginning of each year $t = 1,\ldots, n$) then the accumulated sum of this series of net payments will be equivalent to $\sum_{t=0}^{n-1} b_t(1+i)^{(n-t)}$ at *the end* of year n. This amount is referred to as *the future value* (*FV*) which is defined as:

$$FV = \sum_{t=0}^{n-1} b_t(1+i)^{(n-t)} \tag{A1}$$

The future value *FV* can be converted to a *present value* (NPV, Net Present Value) by calculating the amount that would be equivalent to *FV* at the beginning of year 1. This is done by multiplying *FV* by $(1 + i)^{-n}$ which produces:

$$NPV = \left(\sum_{t=0}^{n-1} b_t(1+i)^{(n-t)}\right)(1+i)^{-n} = \sum_{t=0}^{n-1} b_t(1+i)^{-t} \tag{A2}$$

A one-time amount of A (e.g. an investment at the beginning of year 1) can be converted into an equivalent series of *equally large amounts* over n years by multiplying the one-time amount by the annuity factor $a_{i|n}^{-1}$, where $a_{i|n}^{-1}$ is defined by:

$$a_{i|n}^{-1} \equiv \frac{i}{1-(1+i)^{-n}} \tag{A3}$$

An amount of b calculated as:

$$b = Aa_{i|n}^{-1} = A\frac{i}{1-(1+i)^{-n}} \tag{A4}$$

can, thus, be interpreted as an amount which, if paid out at the *end* of the year t over n years, the series of payments would be equivalent to the one-time amount A at the beginning of year 1.

The formula (A4) can also be used to “even out” a payment series of $b_0,\ldots,$ b_{n-1}, paid out at the beginning of each of a total of n years to an equivalent series of equally large amounts of b, paid out at the end of each of the n years. This is done by multiplying *NPV* in (A2) with the annuity factor $a_{i|n}^{-1}$ which produces:

$$b = a_{i|n}^{-1} NPV = a_{i|n}^{-1} \sum_{t=0}^{n-1} b_t (1+i)^{-t} = \frac{i}{1-(1+i)^{-n}} \sum_{t=0}^{n-1} b_t (1+i)^{-t} \tag{A5}$$

The previous formulas can be used to show that a series of payouts of amount c at the end of each year of a total of n years, is equivalent to a one-time amount (*NPV*), paid out at the beginning of year 1, calculated as:

$$NPV = c a_{n|i} = c \frac{1-(1+i)^{-n}}{i} \tag{A6}$$

Chapter 15
Risk and Uncertainty

15.1 Introduction

Closely linked with the problems associated with the timing of production, as discussed in the last chapter, is risk and uncertainty, which are often associated with the production and sale of products. Implementation of production that takes a very long time is often associated with uncertainty regarding the price which can be obtained for the product by selling at a later date. When it comes to bio-based production such as agricultural crops and other agricultural products, there is also uncertainty as to the production yield, since climatic conditions and diseases may play a role.

The description of how to optimise production in the previous chapters has been based on the assumption that output and output prices are known to the decision maker, i.e. output is a (deterministic) function of the amount of input, and output prices are given by the market (perfect competition) or determined—directly or indirectly—by the decision maker (monopoly and monopolistic competition).

In practice this is often not the case. The relation between input and output may be a *stochastic* function, in the sense that even though one decides the amount of input, the actual amount of output is still uncertain. The quality of the input may be different from what was expected, and therefore there may be unexpected waste. Machines may break down halting production. The labour force may be less productive than anticipated because some may have become sick, or they may have just worked slower than expected. In agriculture there may be a drought, which causes a reduction in crop yields, whilst animals may catch diseases and produce less than predicted. The situation may also be just the opposite in that the quality of the input may be better, machines may last longer, workers may work harder, the weather may turn out to be perfect resulting in higher yields, and animals may remain healthy and grow faster, so that all these factors can also exceed original expectations.

Concerning output prices, it is often difficult to predict what the price will be when the product is ready for sale. A clothes manufacturer producing clothes for

S. Rasmussen, *Production Economics*, DOI: 10.1007/978-3-642-14610-7_15,

customers to buy next summer may have to sell the clothes at much lower prices than expected due to changes in fashion. Farmers growing wheat do not know exactly what the wheat price will be next autumn when the wheat will be ready for sale, whilst the producer of Christmas trees, which have a production period of about 7 years, faces considerable price uncertainty.

We use the concept *uncertainty* to describe situations in which the outcome (here yields or prices) is *not* known with certainty. The drawbacks of uncertainty (*less* output than expected, *lower* output prices than expected) are normally referred to as *risk*. Thus, *uncertainty* is a neutral description of situations which are not certain, whilst *risk* is the concept used to describe the disadvantages of uncertainty.

In the following we will consider some of the methods that can be used for planning and optimisation when output and output prices are uncertain. The simple approach is to use the models we have developed previously and to use *expected* output and *expected* output prices instead of output and output prices. However, although this may work well in many cases, there will also be some cases in which it will not be a good solution, for instance when the firm faces severe losses and the risk of bankruptcy if things go wrong. In these cases, the firm may consider taking out insurance to guard against severe losses. The insurance approach to decision making under uncertainty is treated in the last section of this chapter.

15.2 Expected Profit and Other Criteria

Consider a company facing uncertain profit, either because of uncertain output prices, uncertain yields, or both. Whatever the reason may be, we describe this as a situation in which the firm faces different ***states of nature***, but does not know in advance (i.e. before production is initiated) what the final state of nature will be. For example, consider a farmer who wants to grow either crop A (wheat) *or* crop B (oilseed rape). What will the output price be for the two crops after harvest, when they are to be sold? Will the prices be higher, lower, or the same as last year? What about the climate? Will the growing conditions be good for crop A and bad for crop B, good for crop B and bad for crop A, good for both crops, or bad for both crops? Before production is initiated, the farmer needs to know these things in order to be able to make the right production decision, i.e. to decide which of the two crops to grow, and how much input to apply. However, the farmer cannot know these factors in advance and he is faced with the problem of *decision making under uncertainty*.

To cope with this problem, consider the following simplified example. Assume that there are only two possible price scenarios, "high" or "low", and three different weather conditions, "good", "medium" or "bad" during the growing season. The combination of two possible price scenarios and three different weather conditions makes a total of six ($2 \times 3 = 6$) possible *states of nature*. Assume that there is a 55% chance of "high" prices and a 45% chance of "low" prices. When the price level is "high" the price of crop A is MU 200 and the price of crop B is MU

Table 15.1 A comparison of production A and B

Prices	High			Low			Expected
Weather	Good	Medium	Bad	Good	Medium	Bad	
State of nature	s1	s2	s3	s4	s5	s6	
Probabilities (πs)	0.1375	0.2750	0.1375	0.1125	0.2250	0.1125	
Crop A (MU)							
Total revenue	200,000	120,000	60,000	100,000	60,000	30,000	96,875
Cost	*70,000*	*70,000*	*70,000*	*70,000*	*70,000*	*70,000*	*70,000*
Profit	130,000	50,000	−10,000	30,000	−10,000	−40,000	26,875
Crop B (MU)							
Total revenue	480,000	240,000	0	160,000	80,000	0	168,000
Cost	*110,000*	*110,000*	*110,000*	*110,000*	*110,000*	*110,000*	*110,000*
Profit	370,000	130,000	−110,000	50,000	−30,000	−110,000	58,000

600. When the price level is "low" the price of crop A is MU 100 and the price of crop B is MU 200. There is a 25% chance of "good" weather, a 50% chance of "medium" weather and a 25% chance of "bad" weather. When the weather is "good", crop A yields 10 tons per hectare and crop B yields 8 tons per hectare. When the weather is "medium", crop A yields 6 tons per hectare and crop B yields 4 tons per hectare. When the weather is "bad", crop A yields 3 tons per hectare, and crop B yields 0 tons per hectare. The cost of producing crop A is MU 700 per hectare and MU 1,100 per hectare for crop B regardless of the state of nature. The farmer has a total of 100 hectares. Which crop should the farmer choose?

The probabilities of each of the six states of nature, the total revenue, costs and profit are calculated for each crop and state of nature in Table 15.1. The average (expected) values are shown on the right hand side of the table.[1]

The profit depends on the state of nature, and varies for crop A from MU −40,000 in state s_6 to MU 130,000 in state s_1, and for crop B from MU −110,000 in states s_3 and s_6 to MU 370,000 in state s_1.

The expected price of crop A is $0.55 \times 200 + 0.45 \times 100 =$ MU 155. The expected price of crop B is $0.55 \times 600 + 0.45 \times 200 =$ MU 420. The expected total outcome of crop A is $100 \times (0.25 \times 10 + 0.50 \times 6 + 0.25 \times 3) = 625$ tons, and the expected total outcome of crop B is $100 \times (0.25 \times 8 + 0.50 \times 4 + 0.25 \times 0) = 400$ tons. If we multiply the expected price by the expected outcome we get the expected revenue of MU 96,875 for crop A and MU 168,000 for crop B, as shown in Table 15.1.

One way to make decisions under uncertainty is to choose the crop which gives the highest *expected profit*. Crop B gives an expected profit of MU 58,000, while crop A only gives an expected profit of MU 26,875. Crop B gives the highest profit and is therefore chosen. Notice that we have just substituted prices by *expected prices* and output by *expected output*, and then the calculations have been carried out in the same way as in the earlier chapters when there was no uncertainty.

[1] We assume that prices and weather conditions are independent variables. This may not be the case in real life because the weather influences yield, which may influence prices.

Notice that although crop B gives the highest average profit, crop B also sometimes (probability of 0.25) gives the highest negative profit of MU −110,000. If the firm is able to cope with such a loss, crop B is still the right choice because the average profit is highest. But if a loss of MU 110,000 (with a probability of 0.25) makes the firm go bankrupt, or if other drastic consequences are involved, then the firm will probably be better off choosing crop A instead, because even though crop A gives a lower average (expected) profit, the maximum risk is a loss of MU −40,000, which also only occurs once in every 8–9 years.

Although the obvious choice under risk and uncertainty is to use expected prices and yields, and to choose projects with the highest expected profit, it is not always the best choice. The projects with the highest expected profit may yield the highest income in the long run. But in the long run we are all dead, and the firm may not survive if it takes too many risks.

An alternative way of making decisions under uncertainty is therefore to also focus on the risk of failure, and to choose projects, which give a high expected profit but which also minimise the risk of loss. One such criterion is the so-called *maxi-min criterion*, according to which the decision maker should choose the project with the *highest minimum profit*. In the example above, the project with the highest minimum profit is crop A, which has a higher minimum (−40,000) than project B, which has the lowest minimum (−100,000).

While the maxi-min criterion takes a pessimistic view ("everything probably goes wrong, so I better prepare for the worst case scenario"), the opposite approach is the so-called *maximax criterion*. According to this criterion, the decision maker chooses the project with the highest maximum profit. In the example above, crop B has the highest maximum profit (MU 370,000), and it should therefore be chosen instead of crop A (MU 130,000). The maximax criterion is known as the criterion of the optimistic decision maker—the gambler, who takes risks in the hope of making more money.

There are, of course, other criteria for making decisions under uncertainty. In general, these criteria try to maximise average profits, whilst at the same time ensuring that the risks taken are not too high. How these two factors should be balanced depends on the expectations and the attitude of the individual decision maker, and the financial reserves available to cope with bad times. And because the future outcome is uncertain, there is no objective criterion on which to evaluate the quality of the decision to be taken. If it turns out that the gambler takes a risk which pays off, then we may say that he made the right decision. But if he loses everything, we may also say that he made the wrong decision. But how can a decision be right and wrong at the same time?

15.3 The Cost of Risk

One of the problems with uncertainty is that it involves losses compared to a situation, in which there is no uncertainty. The reason is that decision makers under uncertainty must choose second-best solutions to optimisation problems for which the first-best solution would be a situation, in which the future prices and yields are known when the decision is to be made.

Uncertainty potentially entails costs in two ways: (1) The wrong choice of inputs. (2) The wrong choice of production.

(1) The Wrong Choice of Inputs

This type of uncertainty can be illustrated by an example from agricultural crop production. The farmer does not know how the weather is going to be during the growing season and he therefore has to apply inputs such as seed and fertiliser under uncertainty. For example, if the optimal amount of chemical fertiliser at the start of the growing season depends on the subsequent weather conditions, then there will be uncertainty regarding the correct quantity of fertiliser to apply. If the weather is good during the growing season it would be optimal to apply, e.g. 500 kg. However, if the weather is bad, then it would be optimal to apply, e.g. 300 kg. And finally, if the weather is average, then it would be optimal to apply, e.g. 400 kg. Since the farmer does not know the weather will be in advance, he applies 400 kg of fertiliser each year, which is suitable for average weather conditions (i.e. the second-best solution). If "good", bad and average weather occur with equal frequency, i.e. 1/3, then the strategy to apply 400 kg of fertiliser will mean that the farmer will apply a non-optimal amount of fertiliser in 2 out of 3 years compared to a situation in which he knows what the weather will be like beforehand (first-best solution).

Such losses are difficult to avoid. Improvements in weather forecasting might be an option, although it might just not possible to produce better weather forecasts and even if it is possible, it will involve costs. If the government bears these costs, then society can achieve a welfare improvement if the cost of improving weather forecasting is lower than the financial gains made by farmers through an improved choice of inputs.

Another example is uncertain product prices. If, for example, the price of product A is going to be high after harvesting, then it will be optimal to apply 600 kg fertiliser. If, however, the output price for product A is going to be low, then it will be optimal to apply 200 kg fertiliser. Finally, if the output price is going to be medium, then it will be optimal to apply 400 kg. Since in reality, the farmer cannot know how the price will be in advance, he applies the average of 400 kg fertiliser each year. If high, low and medium prices then occur with equal frequency (i.e. 1/3), then this strategy will mean that the farmer will apply a non-optimal amount of fertiliser in 2 out of 3 years compared to a situation in which he knows what the product price will be beforehand.

Such losses can be avoided if the producer can sign a contract with the buyer beforehand, so that the buyer will pay a fixed price for the product when it is sold. However, a fixed contract price is often lower than the average price that the farmer will receive in the long run, as it is now the buyer who takes over the price risk. The transfer of risks to others normally involves a cost—a kind of insurance premium. The farmer may, therefore, face a financial loss in the long run. In return, he can be certain of the price that he will receive for his crop when it is sold after harvest.

Another possibility to avoid such losses is to select crops for which the optimal supply of inputs is not as sensitive to the weather and prices. If this involves the selection of crops with lower average earnings than more uncertain crops, it means that a certain amount of potential income will be foregone. However, this can be viewed as a kind of insurance premium.

(2) The Wrong Choice of Production

Suppose the farmer can choose between two crops. The first one gives a fixed gross margin of MU 5,000 per hectare. The second gives an average gross margin of MU 7,000 per hectare, but there is considerable uncertainty involved, as the crop gives a gross margin of only MU 1,000 per hectare with a probability of 0.25; a gross margin of MU 13,000 per hectare with a probability of 0.25; and a gross margin of MU 7,000 per hectare with a probability of 0.50. As the farmer cannot afford the low gross margin of MU 1,000, he therefore chooses the crop which gives a fixed gross margin of MU 5,000 per hectare every year.

If the farmer passes on the risk to another party (by taking out an insurance for example), then the choice of the low income crop would involve a total loss equal to MU 2,000 per hectare minus the insurance premium.

The examples above show that although uncertainty typically involves a cost, it may also be possible to reduce, or even to avoid the uncertainty by, for example, signing price contracts, or by choosing crops with a less uncertain yield. And there may even be cases where it is possible to take out insurance with an insurance company.

This opportunity and the possible benefit of transferring uncertainty to others are treated in the formal models in the next section.

15.4 The Expected Utility Model

Empirical observations show that decision makers often prefer a stable/safe income rather than an unstable/unsafe income. This kind of behaviour is called *risk aversion*. At the opposite end of the spectrum are decision makers who only focus on expected profit, and who do not care about the variability of income. Such decision makers are called *risk neutral*.

The preferences of a risk averse decision maker can be illustrated in the form of utility functions similar to the one shown in Fig. 15.1 below.

In Fig. 15.1, income (or rather consumption opportunities) is measured on the horizontal axis and the utility associated with spending income on the vertical axis. The relationship between income and utility is shown as an increasing function, called a von Neumann–Morgenstern (NM) utility function, where utility increases with income but with a decreasing growth rate.

Suppose that a farmer with a NM-utility function, as depicted in Fig. 15.1, has a choice between two projects: (1) an uncertain project A (e.g. the cultivation of a given

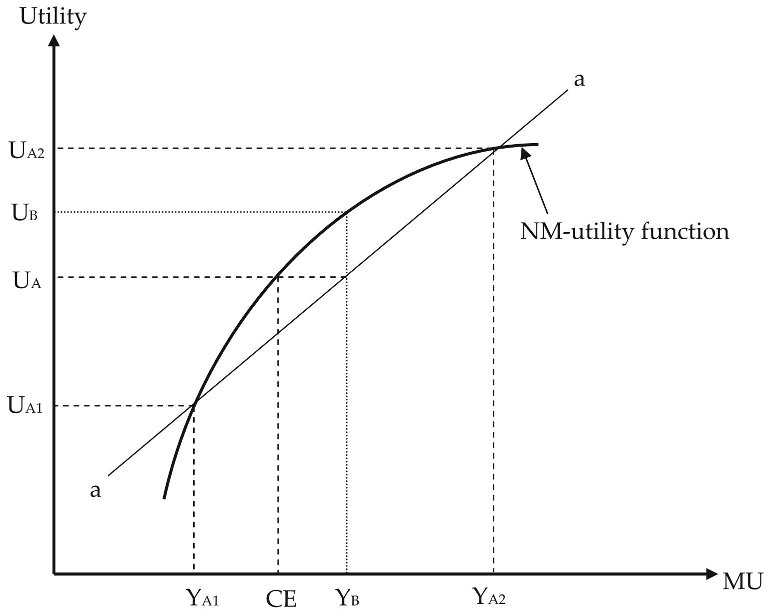


Fig. 15.1 Utility of safe (A) and uncertain (B) project under risk aversion

crop), which provides an income Y_{A1} if it is bad weather, and an income Y_{A2} if the weather is good. "Good" or "bad" weather each have a probability of ½. (2) A safe project B (e.g. wage labour), which in any event (regardless of the weather) provides an income Y_B equivalent to the average earnings of growing crops.

The project A provides a utility of U_{A1} with a probability of ½ (bad weather) and U_{A2} with a probability of ½ (good weather). The expected (average) utility is U_A, calculated as ½ U_{A1} plus ½ U_{A2}. Project B provides a utility of U_B in any case, and this utility is greater U_A. Since project B gives the maximum expected utility, the farmer prefers project B over project A.

Comparing the income of the two projects (the horizontal axis) indicates that both project A and B give the same expected income as equivalent to the Y_B. The expected income (Y_A) of project A is ½ Y_{A1} + ½ Y_{A2} which is precisely equal to Y_B. Although the expected (average) incomes are equal, the project B is preferred, since it gives a higher utility (U_B).

The reason that the risk adverse decision maker chooses project B is that, in contrast to project A, it involves a secure income. The question is how much better is project B than project A. This is also illustrated in Fig. 15.1. If we consider income Y_B then it appears that the farmer also prefers project B even if the safe income is slightly lower. In fact, the secure income from project B may fall below the CE in the figure before the farmer prefers project A (higher utility). The amount of CE is what we call the *certainty equivalent* for the uncertain project A because it is the safe amount of money which gives the same utility as the uncertain project A. The difference between the Y_B and CE is thus the maximum amount the farmer would pay an insurer to insure against the risk associated with project A. This amount is called the risk premium.

While the NM-utility function in Fig. 15.1 illustrates a risk averse decision maker, a *risk neutral* decision maker would have a NM-utility function which is a straight line through the origin. As utility is only a relative measure, we may choose a 45 degree line, in which case the utility is measured in MU and utility maximisation is equivalent to the maximisation of expected income.

The comparison of the two projects A and B may be alternatively illustrated as shown in Fig. 15.2. Figure 15.2 measures the income in case of "bad" weather on the horizontal axis, while income in the event of "good" weather is measured on the vertical axis. The two curves of U_A and U_B are level curves of utility (iso-utility curves), which show the combinations of possible income in the two states of nature that give the farmer the same benefits. When the farmer is risk averse (prefers safe consumption opportunities rather uncertain consumption opportunities, ceteris paribus), then such iso-utility curves have the convex form shown in Fig. 15.2.

It is immediately apparent from the figure that project B gives a higher utility than project A because project B, regardless of the circumstances, gives an income Y_B, which is at a higher iso-utility curve (U_B) than project A, which is on iso-utility curve U_A.

As shown previously, the expected income (Y_A) from project A equals $Y_A = ½ Y_{A1} + ½ Y_{A2}$. If we solve this equation for Y_{A2} we get: $Y_{A2} = Y_A - (½/½) Y_{A1}$, which is a straight line with slope $-½/½ = -1$. This line corresponds to the line

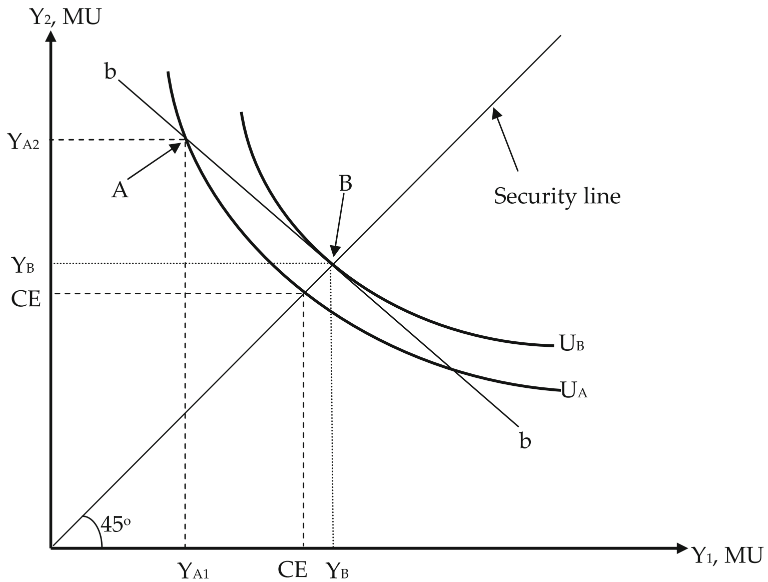


Fig. 15.2 Utility of safe (A) and uncertain (B) projects under risk aversion

bb in Fig. 15.2. The line shows the combinations of income under "bad" and "good" weather conditions, giving an expected income of just Y_A. This line goes through point (Y_B, Y_B), according to the fact that in the example used here, the expected income from project A (Y_A), is equal to the expected income from project B (Y_B).

Based on Fig. 15.2, it is now possible to illustrate the previously mentioned trade-off between expected profit and uncertainty. The objective of this trade-off is to move to the iso-utility curve which provides the highest utility.

15.5 Risk Management and the Insurance Model

As previously mentioned in connection with Fig. 15.1, one can calculate the maximum amount a decision maker is prepared to pay an insurance company to insure against the risk associated with project A. The amount is the difference between the Y_B and CE in Fig. 15.1. The same can be seen in Fig. 15.2, in which the payment of insurance costs on Y_A-CE just reduces the expected income from (Y_B, Y_B) to (CE, CE), where the farmer obtains the same utility (U_A) as in the non-insured project A.

If it is possible to take out insurance at a *fair premium* (*fair* in this context means that the compensation paid to the insurance holder in the long run is equal to the premiums paid by the insurance holder, implying that the insurance holder's expected income is not changed), then the decision maker who chooses project A would be able (through insurance) to change its earnings profile to any point on the line bb in Fig. 15.2. As shown for a risk-averse farmer, it will be optimal to take out *full insurance*. The insurance contract in this example implies that the firm pays an insurance premium of $Y_{A2} - Y_B$ each year, and in exchange receives a compensation of 2 ($Y_{A2} - Y_B$) in those years in which the weather is "bad". Such an insurance means that each year the firm will have a sure (net) income equal to Y_A (= Y_B) with a utility level of U_B available, while the insurance company will have an expected income of zero. At the same time, the farmer obtains a utility increase from U_A to U_B.

An insurance company will not be able to offer decision makers a *fair* insurance if there are costs associated with insurance (transaction costs, etc.). In addition to recovering the claims paid to the firm, the insurer must also cover its costs. This means that the premium will be higher than the fair premium (which is $Y_{A2} - Y_B$).

To illustrate, assume that the insurance cost P is the same regardless of the insured amount and that the insurance company annually collects such costs as lump sums added to the insurance premium. In this case it will still be optimal for the farmer to take out full insurance. When the annual cost P is lower than $Y_B -$ CE (see Fig. 15.2), then the farmer will sign full insurance, with an annual insurance premium of $Y_{A2} - Y_B - P$ in exchange for a compensation of 2 ($Y_{A2} - Y_B$) in the years when the weather is bad. Insurance like this implies that the firm receives a secure income, equivalent to a point on the security line *between* (Y_B, Y_B) and (CE, CE), while the insurance company only covers its costs P. If the insurance cost P is greater than $Y_B - CE$, and the insurer must have its costs covered, then there will be no insurance contract since it would be too expensive for the farmer to insure himself (the incident is not insurable).

If instead (perhaps more realistically) we assume that the insurance costs depend on the size of the insurance and that costs are charged as a proportional addition p to the insurance premium, then the situation is different.

The example here, with an insurance cost of p per MU 1 insurance, involves a trade-off between income in good weather, and income in bad weather of $(1 + p)/(1 - p)$. If, for instance, the insurance cost is 5% of the insured amount ($p = 0.05$), then it is possible to take out insurance along the line with slope $-1.05/0.95 = -1.105$.

This line is drawn as a line cc in Fig. 15.3. As shown, it will still be optimal for the risk averse farmer to take out insurance, since the farmer, by taking out insurance, moves from the starting point A to point C, which provides a slightly higher utility, i.e. U_C. However, in contrast to before it is *not* optimal to take out full insurance. The farmer himself will bear part of the risk and gets an income Y_{C1} in "bad" weather and Y_{C2} in "good" weather.

The larger the costs of insurance, the steeper the line cc in Fig. 15.3. If cc is so steep that it is precisely tangent to the iso-utility curve U_A at point A, then the farmer will no longer have an incentive to take out insurance (it is too expensive) and he will bear the entire risk himself.

The redistribution of income between the states of nature, as illustrated in Figs. 15.2 and 15.3, can be done in ways besides formal insurance as shown here.

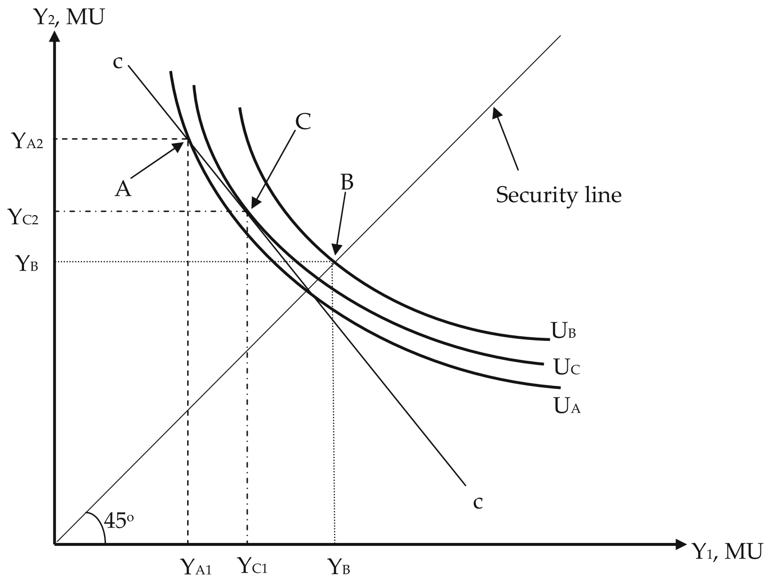


Fig. 15.3 Partial insurance under risk aversion

Saving, borrowing, a change in production to more secure products, price contracts, etc. are examples of such control mechanisms, which can be used to manage risk. As with formal insurances, however, there are often costs associated with such interventions.

The conclusion is that production decisions under uncertainty involve one more dimension than considered in the earlier chapters, i.e. the ***states of nature****. Decision making under uncertainty, therefore, involves choosing control mechanisms that provide a given desired redistribution of income over alternative states of nature at the lowest possible cost.*[2]

[2] Students who wish to study this subject further are referred to the vast literature on this subject, for example Hershleifer and Riley (1992) and Chambers and Quiggin (2000).

Chapter 16
Economic Rent and the Value of Land

16.1 Introduction

The concept of fixed production factors was introduced and discussed in Chap. 12. The main result is that production factors become fixed when it does not pay for the firm to buy more, because the purchase price is higher and the sales price is lower than the internal value of the production factor to the firm. The larger the difference is between purchase and sales price, the higher the probability that a production factor will become a fixed factor.

A production factor earning revenues which are higher than the costs needed to keep it in the present employment is said to earn *economic rent*. The formal *definition of economic rent* is that it is those payments to a production factor that are in excess of the minimum payment necessary to keep it in the present employment. Alternatively, one can define economic rent as being those payments that are in excess of the minimum payment necessary to have the factor supplied.

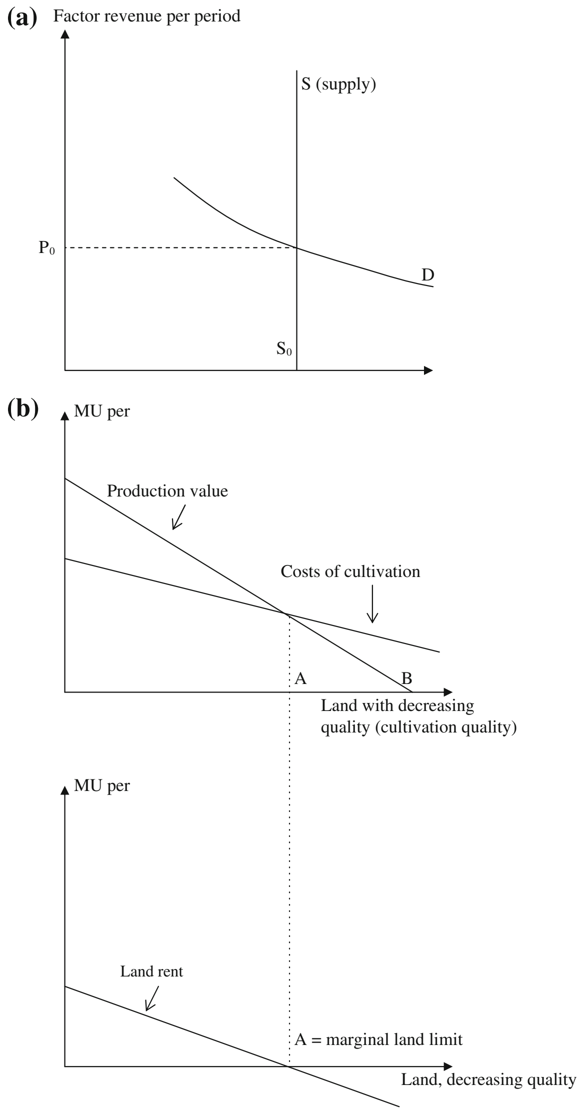
According to the definition, all fixed factors earn some economic rent. However, the original meaning of the concept (*pure* economic rent) is derived from rare and valuable qualities of nature/natural resources such as a rich and productive vineyard, a piece of land where there are raw materials such as oil, copper or gold, a building site with a view of the sea, or a piece of productive agricultural land. The common characteristic is that these factors are unique in the sense that they cannot be reproduced and that they are only available in a certain amount.

Production factors which are available in a certain amount have vertical supply curves, as illustrated in Fig. 16.1. As the supply is fixed at $S_{0,}$ the price is determined by the demand D, and the equilibrium price is P_0. But this price is higher than the minimum payment necessary to have the factor supplied in the amount S_0. In fact, even if the price drops to zero, the supply would still be S_0. Thus, in this case, the fixed production factor earns an economic rent per period, corresponding to the area of the rectangle below the dotted line at $P_{0,}$ and to the left of the supply curve S.

The classical example of this kind of production factor is agricultural land. A country or a region only has a certain amount of agricultural land. Another

S. Rasmussen, *Production Economics*, DOI: 10.1007/978-3-642-14610-7_16,

Fig. 16.1 **a** Supply of fixed production factors
b Definition of land rent



example is taxi licences, the number of which is regulated by the authorities in large cities. Other industries may also be regulated by licences or permits. A person having a natural/unique talent is yet another example. As far as the football club FC Barcelona is concerned, its football players are essential "production factors" and the supply of great football players such as Lionel Messi is certainly limited. At the same time, even if Lionel Messi was only paid one tenth, or even one hundredth, of the huge salary price of he now receives, he would probably still play football. Thus, Messi provides economic rent based on his unique talent.

The interesting question is who receives the economic rent. If the production factor is still supplied in the same amount at a lower price, someone must gain a kind of profit by receiving the high (market) price. It is obvious that in the first place it is the owner of the (unique) production factor who gains the economic rent. It is the owner of land who receives the income (economic rent) $P_0 \times S_0$ from cultivating the land, or who receives the corresponding price when he sells the land. It is the taxi driver who gets the economic rent from driving the taxi or the capitalised value by selling the licence. And it is the football club FC Barcelona which receives the income from the spectators who want to watch Messi playing football.

But this is only part of the story. To see why, consider the situation in which the present owner of the production factor in question has inherited the production factor from his or her parents for example. Thus, assume that the farmer has inherited the land from his parents: that the taxi driver has inherited the licence from her parents: and that Messi in the beginning was just an ordinary Spanish boy who grew up and started playing football in Barcelona. In these cases, the present owner earns all the economic rent. But what about the situation in which the production factor in question is *not* inherited, but instead has to be bought on the open market, so that the farmer has to buy the land, the taxi driver has to buy the taxi licence, and FC Barcelona has to buy Messi from a football club in Argentina. In these cases, the owner pays some, or even all, of the economic rent to acquire the production factor.

The reason is that a production factor which yields an economic rent is valuable to possess. This generates a demand for the production factor, which drives up the price. The farmer, who wants to acquire a certain piece of land, has to pay a price to another land owner in order to obtain it. But how much does he have to pay? This depends on the bargaining power of the buyer and the seller, and no clear answer can be given. But to pay a price which is above the capitalised value of the economic rent would not be profitable.

If the buyer pays a price which corresponds to the capitalised value (i.e. the net present value) of the economic rent, then the buyer transfers all the economic rent to the seller, and he will not therefore gain any economic rent in the end. If, on the other hand, the buyer is able to press the seller to accept a lower price, then the buyer gets some of the economic rent. In general, the distribution of economic rents between the agents or coalition members depends on the relative bargaining power of the involved parties. The distribution of the economic rent generated by Messi between FC Barcelona, the trading agents, and Messi himself (and other agents) thus depends on the individual strength of each. The distribution of land rent between the old farmer who is selling his land, and the young farmer who is going to buy the land also depends on the relative bargaining power of the two.

It is also possible for the government, or other regulating authorities, to extract some, or all, of the economic rent. If, for instance, the government puts a tax on agricultural land, then the demand for land will decrease (the D curve in Fig. 16.1 will move down), and the economic rent left for the agents in the market will therefore be reduced.

In the next part of this chapter we will use the case of agricultural land to further illustrate the concept of economic rent.

16.2 Land Rent and Marginal Land

The concept of economic rent in agriculture (land rent) was originally defined by the British economist Ricardo in 1815. Based on the concept of land rent, the concepts of marginal land and marginal land limit can be defined, and the calculation of the maximum amount to pay for farm land can be derived. In this connection, an analysis of the land price determination in a market occupied by farmers with different cost structures will be presented. Finally, an analysis of the way in which distance can affect the value of land will also be presented.

A description of the land price determination, including a definition of the concept marginal land, is based on Ricardo's land rent theory from 1815.

Land rent is defined as the difference between the production value, when using the land in the best possible way, and the costs of all production factors except land. Hence, land rent is the amount left for the remuneration of the land.

The relationship between production value, costs of cultivation, and land rent is illustrated in Fig. 16.1 above, which shows the calculation of land rent for land of variable quality. The land rent is the difference between the production value and the costs of cultivation. Land rent decreases with declining land quality (site quality) and becomes zero precisely when the costs of cultivation are equal to the production value.

This point (*A*) defines the economic limit of cultivation as it will not be profitable to cultivate land with a poorer quality than at point *A*. The *economic limit of cultivation* A is also referred to as the *marginal land limit*, and land between the marginal land limit and the *cultivation limit B* is simply referred to as *marginal land*, or *extramarginal land*. Hence, marginal land is land that, when used optimally, generates a land rent of zero or less (land between *A* and *B*).

The cultivation limit *B* is solely decided by the quality of the land (site quality/cultivation quality). On the other hand, the economic limit of cultivation *A* depends on site quality, price ratios, and technology, and will typically vary over time. The amount of marginal land can therefore vary over time. Improved price ratios will increase the land rent, implying that the marginal land limit moves towards a poorer land quality (extramarginal acreage will be included for cultivation), while poorer price ratios imply that the land, which was previously cultivated, is now abandoned.

Table 16.1 shows an example of the calculation of land rent for land that is used for growing cereal crops:

In Denmark, the state claims part of the land rent by way of a tax on real property. In this example, the tax on real property is set at MU 400 ha^{-1}, which is deducted in pre-calculated land rent of MU 2,000 ha^{-1}. Deducting the tax on real property produces the land rent of MU 1,600 ha^{-1}, as seen from the point of view of the farmer (personal finance statement).

Table 16.1 Example of the calculation of land rent

Production value, MU per hectare	
Grain, 8,000 kg at MU 1.00	MU 8,000
Straw, 4,000 kg at MU 0.25	MU 1,000
Total	MU 9,000
Unit costs, MU per hectare	
Seeds	MU 500
Fertiliser	MU 1,200
Pesticides	MU 400
Drying, thread	MU 700
Total	MU 2,800
Capacity costs, MU per hectare	
Machinery	MU 3,000
Labour	MU 1,200
Total	MU 4,200
LAND RENT, MU per hectare	MU 2,000
Tax on real property	MU 400
LAND RENT after tax on real property	MU 1,600

The land rent can be used to calculate the maximum rent that a potential tenant must pay to rent the land. If the tenant does not pay taxes on real property (the taxes are paid by the owner), he will pay a maximum of MU 2,000 ha^{-1} to rent the land in question. If he pays precisely this amount, the entire land rent will go to the owner of the land.

16.3 The Use Value of Land

It is also possible to calculate how much should be paid to purchase the land based on the land rent. Presume e.g. that money can be borrowed at the bank at 5% interest per year, and that the life of the loan taken out to purchase the land is infinite (no instalments). What would be the maximum loan that it would be possible to take out to pay for the land?

The unknown loan is referred to as *X*, where *X* is defined by:

$$0.05X \leq \text{MU } 1,600$$

The solution is $X \leq$ MU 32,000. Hence, the maximum amount that should be paid is MU 32,000 ha^{-1}.

This calculation does not consider the possibility of an increase in the value of the land over time, and the related possibility of achieving a value increase profit at a later sale. If the value of land e.g. is expected to increase by 3% a year, the calculation will change to:

$$(0.05 - 0.03)X \leq \text{MU } 1,600$$

which results in a solution of $X \leq$ MU 80,000 ha^{-1}.

The method used here to calculate the (use) value of the land when using land rent and borrowing rate (market rate) can be represented in the following simple formula:

$$\text{Land value} = \frac{\text{Landrent}}{i_j} \tag{16.1}$$

in which the interest rate i_j is the market rate adjusted for the expected price increase of land value (the market rate minus the expected price increase of land).

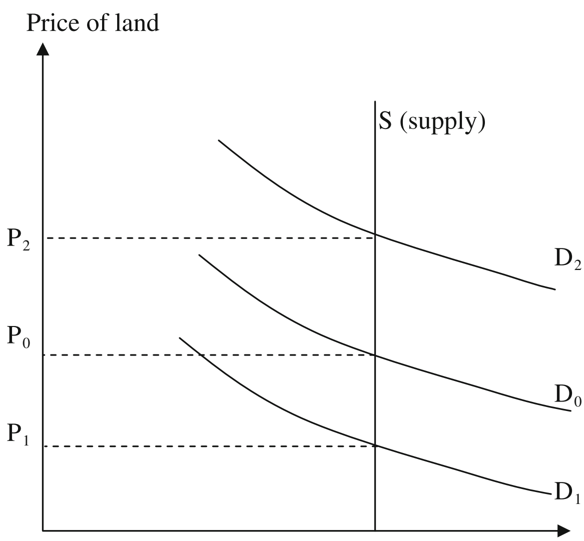
The calculation shown only produces the so-called *use value,* which is the value of the land according to the possible *operational usage* of the land. The price, at which the land is traded, in practice, normally depends on many other factors besides the use value, as a number of non-economic factors will also have an influence. Intangible values, such as e.g. location, the condition of the farmhouse, if there are good hunting areas etc., will naturally also have an influence on the actual market price, as does the general competition in the market. However, the use value of the land, as described above, provides a good basis for a potential buyer to estimate a start offer for land in connection with a deal.

16.4 The Land Price Determination

We will now examine the relationship between the land rent mentioned above and the land price determination.

The general macro-economic description of the land price determination is illustrated in Fig. 16.2 below.

Fig. 16.2 Land price determination



Land is present in a given, fixed supply illustrated by the vertical supply curve S in Fig. 16.2. Three different demand curves D_0, D_1, and D_2, and the related corresponding equilibrium prices given by P_0, P_1, and P_2, have been plotted on the figure.

But how are the illustrated demand curves produced?

Presume e.g. that the total supply of land is 280 hectares, and that there are three farmers wanting land in the market. The three farmers are different in the sense that the "skilful" farmer (A) is capable of achieving a high profit (land rent), the "average" farmer (B) only achieves an average profit (land rent), whilst the "poor" farmer (C) achieves a low profit (land rent). For the sake of simplicity, the profit (land rent) per hectare is presumed to be constant, regardless of the number of hectares.

It is, furthermore, presumed that a law has been introduced which stipulates that farmers cannot own and run farms larger than 100 hectares. In economic terms therefore, the price of land of more than 100 hectares is infinitely high.

The economic situation for the three farmers can be illustrated as shown in Fig. 16.3.

The profit (land rent) per hectare is constant. For the "skilful" farmer it is MU a per hectare, for the "average" farmer it is MU b per hectare, and for the "poor" farmer it is MU c per hectare.

The market demand curve can now be derived from the three land rent curves in Fig. 16.3, as the land rent is the amount that a farmer would be willing to pay to rent the land. The corresponding land value (the maximum purchase price) is calculated as in (16.1). If the land values corresponding to a, b, and c are referred to as A, B, and C (where e.g. $A = a/i$, and i is the interest rate), it results in the following graphical illustration Fig. 16.4 of the determination for this small illustrative example:

Hence, the "skilful" farmer will be willing to pay up to MU A per hectare, the "average" farmer up to MU B per hectare, and the "poor" farmer up to MU C per hectare. The demand curve D is produced by the decreasing, staircase-shaped line shown in Fig. 16.4, corresponding to the D demand curves in Fig. 16.2.

Presuming that the quality of the 280 hectares of land is the same, and that there is no possibility whatsoever of differentiating between the buyers, then the equilibrium price will be MU C per hectare.

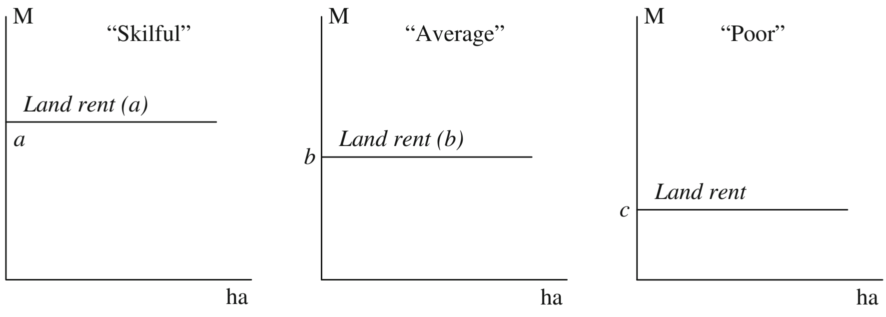


Fig. 16.3 Earnings per hectare for three different farmers

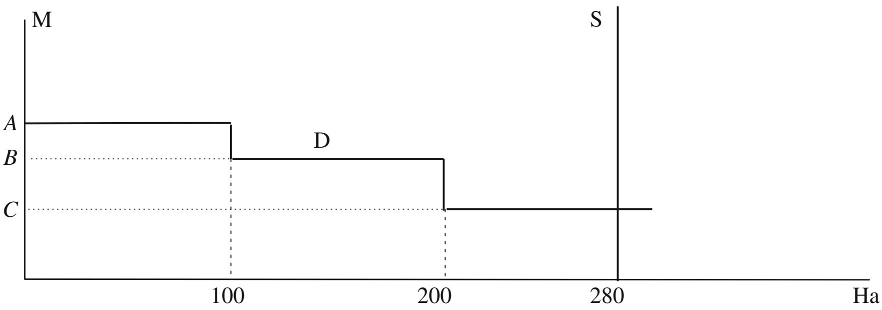


Fig. 16.4 Land price determination (*S* supply, *D* demand)

The "skilful" farmer who, with 100 hectares, has a total land rent of 100*A* would, with a price of *C*, be able to keep part of the land rent, i.e. $100(A-C)$. The "average" farmer who, with 100 hectares, has a total land rent of 100*B* would, with a price of *C*, also be able to keep part of the land rent, i.e. 100(*B*–*C*). The "poor" farmer would, with a price of MU *C* per hectare, have to pay all his/her land rent 80*C* for the 80 hectares and would, therefore, not be able to keep any of the land rent.

Example 16.1: The "skilful" farmer has a land rent of $a =$ MU 1,600 ha^{-1}. The "average" farmer has a land rent of $b =$ MU 1,000 ha^{-1}. The "poor" farmer has a land rent of $c =$ MU 400 ha^{-1}. With an interest rate *i* of 4% year^{-1} (market rate minus the expected price increase of land), the prices *A*, *B*, and *C* in Fig. 16.4. will be equal to MU 40,000, MU 25,000, and 10,000 ha^{-1}, respectively.

16.5 Location and Land Rent

The traditional model for the price determination of land located at different distances from a market (an isolated city) was developed by von Thünen back in the 1800s. A brief example illustrates the basic principle behind von Thünen's model.

Farmer A has a distance of 100 km to the market and receives MU 4.80 kg^{-1} milk. Farmer B has a distance of 500 km to the market and receives MU 4.00 kg^{-1} milk, as the milk processer deducts MU 0.80 kg^{-1} milk to cover the extra costs of transportation.

Farmer A and B have similar cost curves. For farmer B, the price of MU 4.00 kg^{-1} provides for exact cost coverage as shown in Fig. 16.5. As farmer A is paid a higher price than farmer B, farmer A produces more than farmer B, and furthermore receives an economic rent (location rent) which is equal to the difference between the product price of MU 4.80 and the average cost of a little more than MU 4.00 kg^{-1} (see Fig. 16.5).

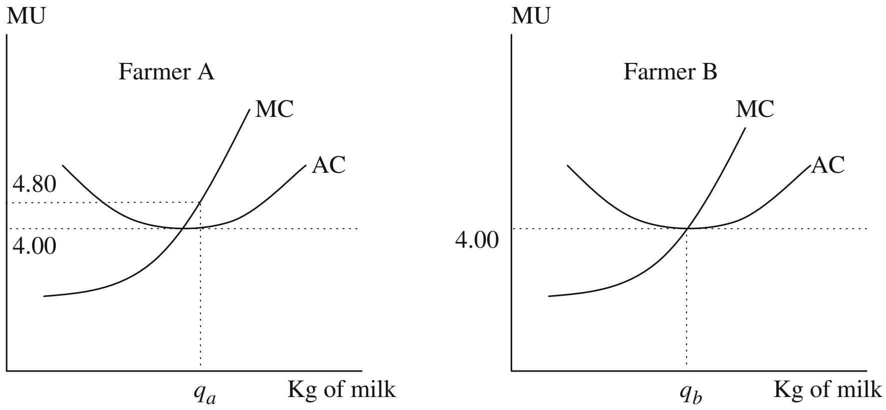


Fig. 16.5 Economic rent achieved by higher product price

If farmer B is rational, he/she will—everything else being equal—also want to move his/her production closer to the market to achieve a positive location rate of MU 0.80 kg^{-1} milk, like farmer A. Farmer B will be willing to pay a higher price for land that is located closer to the market. Farmer B will in fact be willing to pay an excess price for the land of up to $0.80q_b$ divided by the interest rate i. In this way, the prices offered for land increase. This phenomenon is called *capitalisation*. For new producers who buy land at this higher price, up to the entire land rent will be consumed by increasing costs due to the high price paid for the land. And if the price paid for the land means that all the land rent is consumed due to the higher purchase price, then the farmer will in fact not be better of economically by being closer to the market.

Chapter 17
Production of Multiple Products

17.1 Introduction

A company often produces several products, which gives rise to new issues compared to the situation in which only one product is being produced. If, for example, there are limited amounts of input at disposal, then how this input should be allocated between the various products needs to be decided. Multiple products can also give rise to economic issues concerning production in another way. It may be the case that the production of the multiple products results in advantages, e.g. a kind of synergy. Producing two products at the same company can also give rise to advantages regarding cost savings when compared to separate production.

17.2 The Product Transformation Curve

The analysis will be based on the assumption that the company has an input x available in a limited fixed amount x^0, and that the company can produce two outputs y_1 and y_2 by using this input. The question is how to allocate the limited amount of fixed input when producing the two outputs. An example could be the use of a given acreage of farm land for growing wheat or barley. Another example could be a given amount of labour which can be used to produce milk or oilseed rape.

The issue presupposes that the amount of x constitutes a real limitation. If x could simply be purchased in arbitrary amounts, there would not be any particular problem in connection with the production of the two products as the optimisation of the production could be carried out for each product individually. However, there are usually fixed inputs which constitute a real (physical) limitation on production. In addition to this, there can be synergistic effects between the two products. This last issue will be discussed later.

Whether x^0 is considered a given amount of an individual input or a basket of multiple inputs is of no importance. Substitution between multiple inputs has been

S. Rasmussen, *Production Economics*, DOI: 10.1007/978-3-642-14610-7_17,

discussed above (see Chap. 4), and including the possibility of substitution between multiple inputs does not add anything to the discussion here. In the following, x is simply considered a fixed input, either in the form of an individual input (x is a scalar) or as a basket (a vector) of inputs.

Consider production functions for each of the two products. Figure 17.1 shows examples of such production functions with one input–one output.

The left and middle columns of Fig. 17.1 each show three examples of production functions for y_1 and y_2. The right column of the figure shows the combinations of y_1 and y_2 that could be produced with the input amount x^0. The curves with bold lines (A, B and C) are the so-called *product transformation curves* or simply *transformation curves*. Such curves are also referred to as *production possibility curves*.

It can be seen that, with normal production functions (production functions with diminishing marginal returns, as in the centre row in Fig. 17.1); the transformation curve is a concave curve (with the indention towards the zero point). The transformation curves will therefore be drawn in this way in the following.

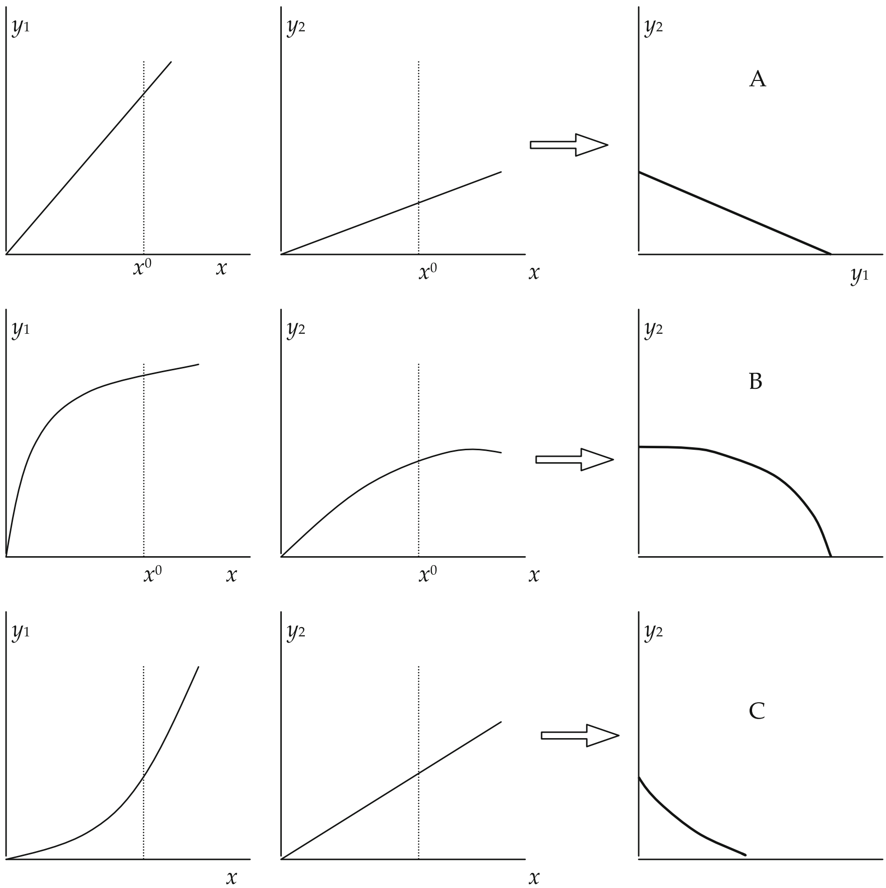


Fig. 17.1 Production functions (*left* and *centre*) and transformation curves (*right*)

A transformation curve can be mathematically described based on an “inverted” production function, thus:

$$x = g(y_1, y_2) \tag{17.1}$$

in which the function g describes how much of input x is used as a function of the production of y_1 and y_2.

To describe the relationship between y_1 and y_2, the total differential can be calculated as:

$$dx = \frac{\partial g}{\partial y_1} dy_1 + \frac{\partial g}{\partial y_2} dy_2 \tag{17.2}$$

The transformation curve is thus defined by the combinations of y_1 and y_2 that can be produced with the given amount x. This can be expressed in the way that changes in the production of y_1 and y_2 must take place in such a way that the use of x does not change, i.e. that $dx = 0$. If 0 is inserted in the left hand side of (17.2), the transformation curve can thus be described as:

$$0 = \frac{\partial g}{\partial y_1} dy_1 + \frac{\partial g}{\partial y_2} dy_2 \tag{17.3}$$

and if the following is used:

$$\frac{\partial g}{\partial y} = \frac{1}{\frac{\partial y}{\partial g}} = \frac{1}{\frac{\partial y}{\partial x}} = \frac{1}{MPP} \tag{17.4}$$

the slope of the transformation curve can be expressed as:

$$\frac{dy_2}{dy_1} = -\frac{MPP_{i2}}{MPP_{i1}} \tag{17.5}$$

in which MPP_{ij} is used as the expression for the marginal product when using input x_i for the production of output y_j.

The expression in (17.5) is referred to as *the rate of product transformation*, or simply as *the transformation rate* (RPT_{21}). This expresses how much the production of y_2 must be changed, if you want to produce more of y_1 but only have a given amount of input x at your disposal.

While the normal transformation curve has a shape as shown in B in Fig. 17.1, it is possible to establish other relationships in the production of two outputs with a given input amount. Figure 17.2 shows four classical examples of such relationships.

The relationship between the two products is said to be *competitive* if the transformation curve has a negative slope ($dy_2/dy_1 < 0$).

The relationship between the two products is said to be *supplementary* if it is possible to increase the production of one output for a given input amount without this affecting the production of the other output ($dy_2/dy_1 = 0$). An example from agriculture is the season-related variation in the need for labour. If the production is based on fixed family labour, there will be a possibility of starting/increasing

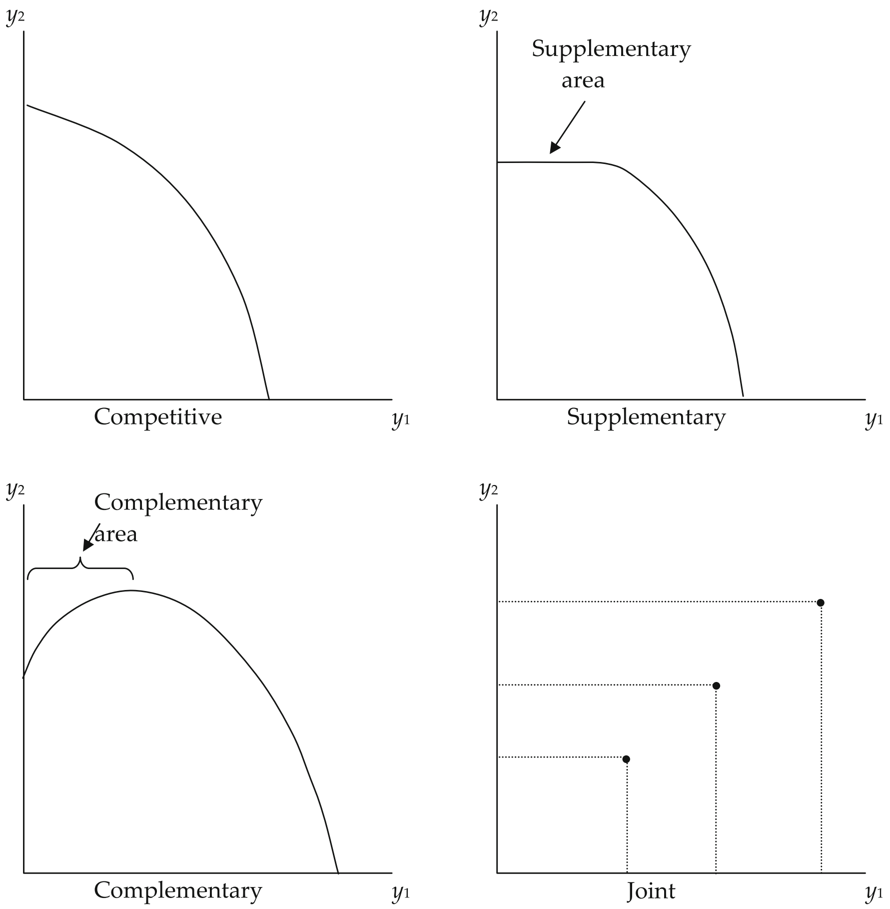


Fig. 17.2 Competitive, supplementary, complementary, and joint productions

another production in periods with a limited need for labour (e.g. at crop farms during the winter season) without having to reduce the first production.

The relationship between the two products is said to be *complementary,* if it is possible to increase the production of one output for a given input amount, while at the same time increasing the production of another output ($dy_2/dy_1 > 0$). An example from agriculture is the effect of *crop rotation* on crop production when the crop order can affect the yield. If only one crop is grown on the entire acreage, it may be possible to increase the yield by introducing another crop in the crop rotation. If this is the case, it is referred to as complementary production.

Complementary relationships between different productions affect the production costs. If the unit costs in a production can be reduced by combining the production with another production, it is referred to as *economies of scope*. An example of this is the production of livestock (y_1) under conditions where the disposal of livestock manure involves considerable costs, and the production of sales crops (y_2) is based on the purchase of expensive fertiliser. If these two productions are combined, so that the livestock manure is used as fertiliser in the production of sales crop resulting in cost savings, it is referred to as *economies of scope* in production.

If a product cannot be produced without the production of another product, it is referred to as *joint production*. The relationship between the two outputs is, hence, given by the discrete points as shown in the bottom figure to the right in Fig. 17.2. Classical examples from agriculture include the production of beef and hide, mutton and wool and milk and meat. It is sometimes possible to vary the combination of the two outputs, so that it will be a case of substitution within a certain framework instead of discrete points. The relationship between milk and beef can, thus, vary according to the choice of cattle breed.

Example 17.1 The mathematical relationships can be further illustrated by a simple example.

Product y_1 is produced by using an input x_1 according to the production function:

$$y_1 = 2x_{11} \tag{17.6}$$

while product y_2 is produced by using input x_1 according to the production function:

$$y_2 = 3x_{12} \tag{17.7}$$

in which x_{ij} is the use of input i in the production of product j.

The total use of x_1 is a function of y_1 and y_2 given by:

$$x_1 = \frac{1}{2}y_1 + \frac{1}{3}y_2 \tag{17.8}$$

which results in the following total differential:

$$dx_1 = \frac{1}{2}dy_1 + \frac{1}{3}dy_2 \tag{17.9}$$

Setting dx_1 equal to zero produces:

$$\frac{dy_2}{dy_1} = -\frac{MPP_{i2}}{MPP_{i1}} = -\frac{\frac{1}{2}}{\frac{1}{3}} = -\frac{3}{2} \tag{17.10}$$

Hence, the product transformation rate is equal to −3/2, and the production possibility curve is therefore equal to a straight line with the slope −3/2. The two productions y_1 and y_2 are competitive.

17.3 The Optimal Combination of Output

Transformation curves, as illustrated in Figs. 17.1 and 17.2, are based on a given available amount of input $x = x^0$. There will, therefore, be different

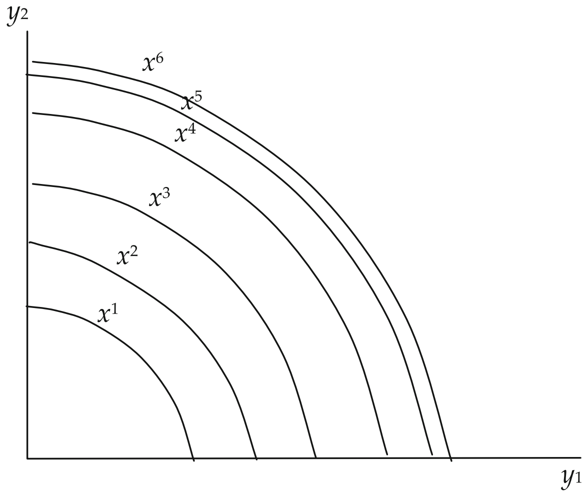


Fig. 17.3 Multiple transformation curves

transformation curves corresponding to different input amounts. This is illustrated in Fig. 17.3.

For each amount of input $x^1, \ldots, x^6$, there is a transformation curve showing the combinations of y_1 and y_2 that could be produced with the given input amount. The curves have been drawn under the assumption that $x^1 <, \ldots, < x^6$, so that a larger amount of input is a possibility for a larger production. With diminishing marginal returns, the increase in production by adding increasing amounts of input x will gradually decline. In the figure, this is illustrated by the transformation curves being closer together with the addition of increasing amounts of x. At some point this will result in maximum production, and the transformation curves for still increasing amounts of x will coincide (not illustrated).

The optimisation problem in connection with a given amount of input consists of determining the combination of y_1 and y_2 that results in the highest total revenue. The revenue R is calculated as:

$$R = p_1 y_1 + p_2 y_2 \tag{17.11}$$

which entails a relationship between the produced amounts of y_1 and y_2, and the related corresponding revenue is given by:

$$y_2 = \frac{R}{p_2} - \frac{p_1}{p_2} y_1 \tag{17.12}$$

which is a straight line (referred to as *the iso-income line*) with an intersection point on the vertical axis at the point R/p_2, and with a slope equal to $-p_1/p_2$.

The line is illustrated by the dotted line in Fig. 17.4. It can be illustrated graphically that the highest value of the revenue R is achieved by producing q_1 units of y_1 and q_2 units of y_2, which is exactly what happens to the combination at the point where *the iso-income line* is tangent to the production possibility curve.

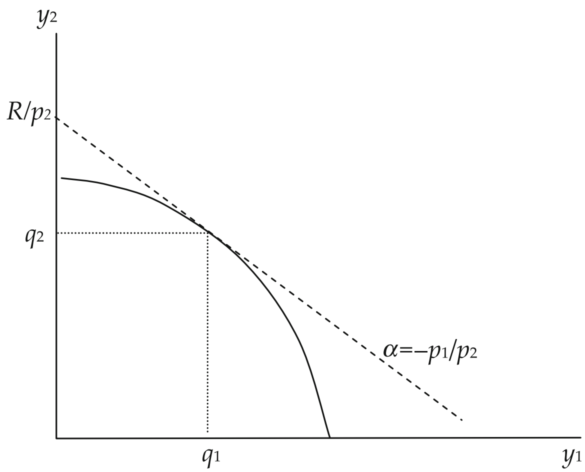


Fig. 17.4 Optimal combinations of y_1 and y_2

With this combination, the intersection point R/p_2 is shifted upwards along the vertical axis to its highest possible position.

As previously shown (see (17.5)), the slope of the production possibility curve is $-MPP_{i2}/MPP_{i1}$. For an optimal combination of y_1 and y_2, the slope of the iso-income line is precisely equal to the slope of the production possibility curve. An optimal combination of y_1 and y_2 is, therefore, characterised by the following condition:

$$\frac{dy_2}{dy_1} = -\frac{MPP_{i2}}{MPP_{i1}} = -\frac{p_1}{p_2} \tag{17.13}$$

This criterion can also be derived mathematically by use of the Lagrange method with the formulation of the following optimisation problem:

$$\text{Max}\{R\}, \quad \text{where } R = p_1y_1 + p_2y_2 \tag{17.14a}$$

under the constraint that:

$$x = g(y_1, y_2) \tag{17.14b}$$

The Lagrange function L is expressed as:

$$L = p_1y_1 + p_2y_2 + \theta(x - g(y_1, y_2)) \tag{17.15}$$

in which θ is the Lagrange multiplier.

The optimum is found by differentiating L with regard to the three variables x_1, x_2, and θ, and setting the derivatives equal to zero. The solution is as follows:

$$p_1 - \theta\frac{\partial g}{\partial y_1} = 0 \tag{17.16a}$$

$$p_2 - \theta\frac{\partial g}{\partial y_2} = 0 \tag{17.16b}$$

$$x - g(y_1, y_2) = 0 \tag{17.16c}$$

If (17.4) is used, the conditions (17.16a) and (17.16b) can be combined to form the following condition for optimality:

$$\frac{p_1}{p_2} = \frac{MPP_{i2}}{MPP_{i1}} \tag{17.17}$$

and if (17.5) is used (17.17) is found to be in line with (17.13) which was found based on the graphical solution.

The conditions (17.16a) and (17.16b) can also be expressed as:

$$\frac{p_1}{\frac{\partial g}{\partial y_1}} = \theta \tag{17.18a}$$

$$\frac{p_2}{\frac{\partial g}{\partial y_2}} = \theta \tag{17.18b}$$

If (17.4) is used again (17.18a) and (17.18b) can be written as:

$$VMP_{i1} = VMP_{i2} = \theta \tag{17.19}$$

in which VMP_{ij} is the value of the marginal product (p_jMPP_{ij}) when using input x_1 for the production of output y_j.

The condition (17.19) determines that a given amount of input x_i should be distributed between the two productions of y_1 and y_2, respectively, so that the value of the last input unit (*VMP*) is the same in both productions.

As can be seen from (17.19), the value of this last unit (the marginal value) is exactly equal to the Lagrange multiplier θ. This in fact corresponds to the interpretation of the Lagrange multiplier in the Lagrange function (17.15), as the Lagrange multiplier is equal to the shadow price of the restriction.

The condition (17.19) can be generalised for more than two products. Hence, the following is true for a total of k outputs:

$$VMP_{i1} = \ldots = VMP_{ik} = \theta \tag{17.20}$$

Example 17.2 A farmer is unsure about how many hours to use in the production of (1) blackcurrants and (2) pigs. He/she also has the option of (3) working outside the home which pays an hourly wage of MU 150.

According to (17.20), labour should be distributed between the three productions in the way that:

$$VMP_{i1} = VMP_{i2} = VMP_{i3} \tag{17.21}$$

As VMP_{i3} is known (=MU 150), the answer is that the labour should be distributed so that an extra hour of work would pay the same, namely MU 150, in the blackcurrant and pig production.

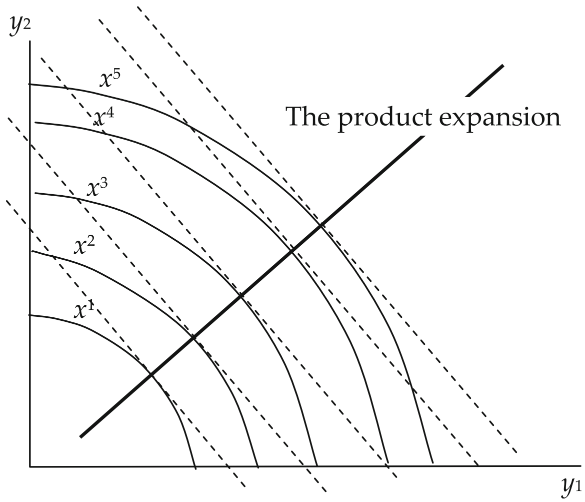


Fig. 17.5 The product expansion path

17.4 Product Expansion Path

Figure 17.4 contained an illustration of how a given input amount can be used to determine the optimal combination of two productions. What if more input becomes available? How should this extra input be distributed between the two productions.

The relationship is described in Fig. 17.5. Initially, the available amount of input is x^1, and the optimal production is found when the iso-income line (the dotted lines) with the slope $-p_1/p_2$ is tangent to the production possibility curve.

With increasing amounts of input, the production possibilities, illustrated by transformation curves that are placed higher and higher, increase. For each input level, the price line is drawn tangent to the transformation curve–tangent points illustrating the optimal production combination. Connecting these tangent points produces the curve drawn with a bold line, *the product expansion path*.

The product expansion path indicates by which "path" production should be expanded if more input becomes available. Figure 17.5 shows a straight line where the two productions are expanded at a constant ratio. This might not be the case. Depending on the production technology, the product expansion path can take other forms, so that the optimal expansion is not necessarily to be found at a proportional increase.

17.5 Intermediate Products and Trade

Companies sometimes produce products which are not sold but are instead used as inputs to other production processes. A typical example from agriculture is the growing of roughage as fodder for cattle.

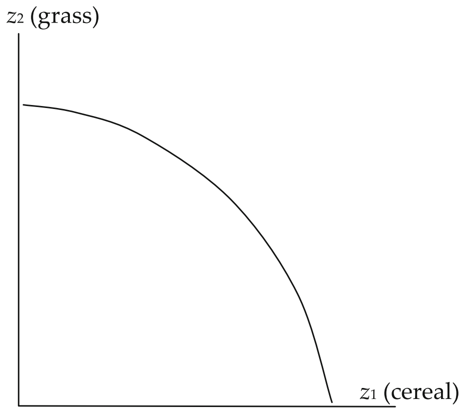


Fig. 17.6 The product possibility curve

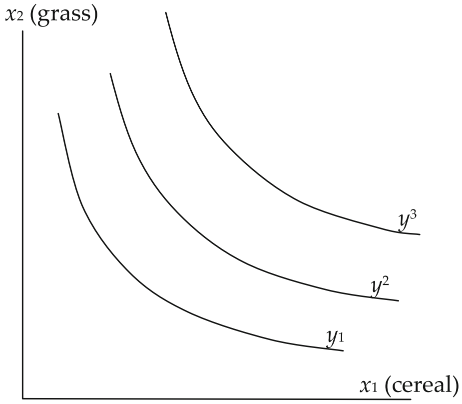


Fig. 17.7 Isoquants

An example: A farmer grows two products grass (z_1) and cereal (z_2). With the given acreage x at his disposal, the production possibility curve is shaped as shown in Fig. 17.6.

The farmer uses the two (intermediate) products as fodder for the cattle. The isoquants for the production of different product amounts (beef) (y^1, y^2, y^3) is shown in Fig. 17.7.

The farmer's aim is to produce as much beef as possible. How is the production of cereal and grass optimised?

Graphically, the solution is relatively simple. The optimal combination is found when the production possibility curve is tangent to an isoquant. This can be seen in Fig. 17.8 below in which the two curve sets in Figs. 17.6 and 17.7 have been combined.

The optimal production of cereal and grass is q_1 and q_2, respectively. For these amounts, the slope of the production possibility curve ($-MPP_{i2}/MPP_{i1}$) is exactly equal to the scope of the isoquant ($-MPP_1/MPP_2$). Hence, the optimal combination for the production of intermediate products is found when:

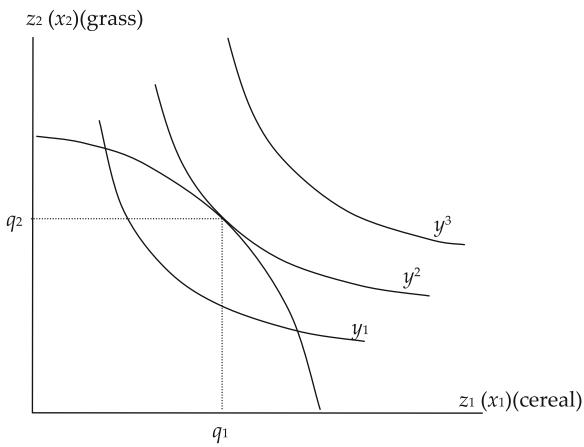


Fig. 17.8 The optimisation of the production of intermediate products

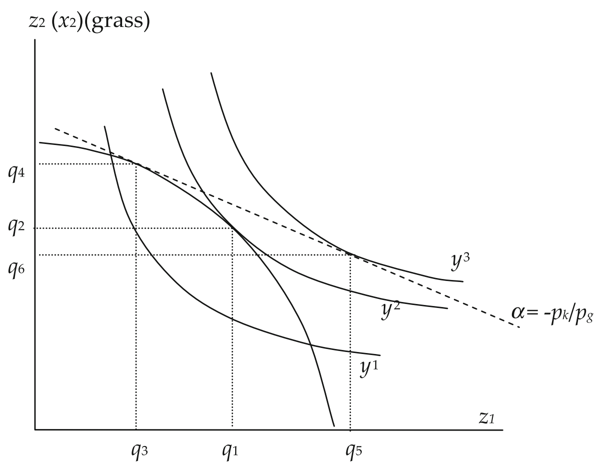


Fig. 17.9 The optimisation of the production of intermediate products with trade

$$\frac{MPP_{i2}}{MPP_{i1}} = \frac{MPP_1}{MPP_2} \tag{17.22}$$

The model in Fig. 17.8 entails that the producer produces exactly the amount that is consumed, i.e. a production and consumption equilibrium. However, it will often be possible to *trade intermediate products*. It might be an idea to buy in order to supplement one's own production, or to sell the amount that cannot be used in one's own production.

In situations in which it is possible to trade, the situation changes compared to Fig. 17.8. The possibility of trade can be illustrated as shown in Fig. 17.9.

The possibility of trade is assumed to be present when cereal can be purchased and sold at the price p_k, and grass can be purchased and sold at the price p_g. This is illustrated by the dotted price line with the slope $-p_k/p_g$ in Fig. 17.9.

With this price ratio, it is optimal (compared to the pre-trade situation, Fig. 17.8) to reduce the production of cereal from q_1 to q_3 and, on the other hand, increase the production of grass from q_2 to q_4.

Hence, the optimal production is (q_3, q_4). However, the optimal *consumption* is different from the optimal *production* as, apart from changing production, trade will also take place. The farmer sells part of the grass, namely q_4 minus q_6. On the other hand, he/she purchases cereal in the amount of q_5 minus q_3. Hence, the final amount for consumption as fodder is q_5 units of cereal and q_6 units of grass.

As can be seen, the production has increased. It used to be y_2, but is now y_3. This improvement in the producer's situation is achieved solely through trade.

17.6 Multiple Inputs and Outputs

To conclude this chapter, the criteria for optimal production in connection with multiple inputs and outputs are derived. To start with, the problem is simplified so that it only concerns two (variable) inputs and two outputs. The result of this can, however, be generalised to include more than two inputs and outputs.

The company is presumed to be able to produce two outputs, y_1 and y_2, with the use of two (variable) inputs, x_1 and x_2. The production of y_1 is:

$$y_1 = f_1(x_{11}, x_{21}) \tag{17.23}$$

and the production of y_2 is:

$$y_2 = f_2(x_{12}, x_{22}) \tag{17.24}$$

The budget is, furthermore, presumed to be restricted to MU C^0.

The optimisation problem can thus be formulated as:

$$\text{Max}\{p_1 f_1(x_{11}, x_{21}) + p_2 f_2(x_{12}, x_{22})\} \tag{17.25a}$$

under the constraint:

$$w_1(x_{11} + x_{12}) + w_2(x_{21} + x_{22}) = C^0 \tag{17.25b}$$

in which x_{ij} is the use of input i for the production of product j.

The Lagrange function thus equals:

$$L = p_1 f_1(x_{11}, x_{21}) + p_2 f_2(x_{12}, x_{22}) + \lambda(C^0 - w_1(x_{11} + x_{12}) - w_2(x_{21} + x_{22}))$$

Differentiating L with regard to the five variables and setting the derivatives equal to zero will result in the following necessary condition for an optimal production:

$$\frac{\partial L}{\partial x_{11}} = p_1 MPP_{11} - \lambda w_1 = 0 \tag{17.26a}$$

$$\frac{\partial L}{\partial x_{12}} = p_2 MPP_{12} - \lambda w_1 = 0 \tag{17.26b}$$

$$\frac{\partial L}{\partial x_{21}} = p_1 MPP_{21} - \lambda w_2 = 0 \tag{17.26c}$$

$$\frac{\partial L}{\partial x_{22}} = p_2 MPP_{22} - \lambda w_2 = 0 \tag{17.26d}$$

$$w_1(x_{11} + x_{12}) + w_2(x_{21} + x_{22}) = C^0 \tag{17.26e}$$

Solving (17.26a–17.26d) with regard to λ and setting these four expressions of λ equal to each other produces the following condition for optimal production:

$$\frac{VMP_{11}}{w_1} = \frac{VMP_{12}}{w_1} = \frac{VMP_{21}}{w_2} = \frac{VMP_{22}}{w_2} = \lambda \tag{17.27}$$

The condition in (17.27) expresses that for each input and for each product, the ratio between the value of the marginal product and the input price is equal to the factor λ for an optimal production. Hence, the ratio between the value of the marginal product and the input price must be the same, regardless of the combination of input and output in question.

The Lagrange multiplier λ expresses the shadow price of the budget constraint used as the constraint in (17.25a, 17.25b). Hence, λ expresses the extra profit that could have been achieved by having one more MU at disposal. Hence, in situations when the budget constraint is effective (the producer would have liked to have had a bigger budget), λ is larger than 1, and the ratio between the value of the marginal product and the input price is, thus, according to (17.27) also larger than 1. If the budget is increased, the shadow price is reduced, and in the borderline case when λ is equal to 1, the budget constraint is no longer effective, and the producer can achieve the profit maximum, which is where the ratio between the value of the marginal product and the input price is equal to 1.

Example 17.3 A gardener has MU 100 at his/her disposal (budget constraint). With this amount he/she can buy two inputs, x_1 and x_2, that can be used for the production of the product y_1 or y_2 (or both). The production function for y_1 is $y_1 = f_1(x_{11}, x_{21}) = 6x_{11}^{0,3}x_{21}^{0,5}$, and the production function for y_2 is $y_2 = f_2(x_{12}, x_{22}) = 9x_{12} + 6x_{22} - x_{12}x_{22}$, in which x_{ij} is the use of input i for the production of output j. The price of x_1 is MU 8 ($w_1 = 8$) and the price of x_2 (w_2) is MU 12. The price of the product y_1 is MU 7, and the price of the product y_2 is MU 6.

The Lagrange function L is expressed as:

$$L = 7 \times 6x_{11}^{0,3}x_{21}^{0,5} + 6 \times (9x_{12} + 6x_{22} - x_{12}x_{22}) \\ + \lambda(100 - 8(x_{11} + x_{12}) - 12(x_{21} + x_{22}))$$

If a differentiation is carried out with regard to the five variables, the following conditions for optimal production are generated:

(1) $7(1,\ 8x_{11}^{-0,7}x_{21}^{0,5}) = 8\lambda$

(2) $6(9 - x_{22}) = 8\lambda$

(3) $7(3x_{11}^{0,3}x_{21}^{-0,5}) = 12\lambda$

(4) $6(9 - x_{12}) = 12\lambda$

(5) $8(x_{11} + x_{12}) + 12(x_{21} + x_{22}) = 100.$

Dividing (1) by (3) produces:

(6) $$\frac{1,8x_{21}}{3x_{11}} = \frac{8}{12}$$

Dividing (2) by (4) produces:

(7) $$\frac{9 - x_{22}}{6 - x_{12}} = \frac{8}{12}$$

Finally, dividing (2) by (3) produces:

(8) $$\frac{6(9 - x_{22})}{3x_{11}^{0,3}x_{21}^{-0,5}} = \frac{8}{12}$$

The conditions (5), (6), (7), and (8) now constitute a set of four equations with four variables x_{11}, x_{21}, x_{12}, and x_{22}. The reader is encouraged to solve the equations. Then λ can be calculated.

Chapter 18
The Linear Programming Model

18.1 Introduction

The linear programming model which is described in the following is based on a special type of production function, the so-called *Leontief production function*.

The Leontief production function is different from the general neoclassical production function in that it consists of straight lines and the isoquants are L-shaped, indicating that it is *not* possible to *substitute* between input factors, and that the input factors, therefore, are used at a fixed ratio.

In the neoclassical production model, a given product amount can be produced with different combinations of input factors.

In case of e.g. two input factors (input), z_1 and z_2, the isoquant in Fig. 18.1a shows that the product amount y^0 can be produced with different combinations of z_1 and z_2—i.e. that it is possible substitute z_1 for z_2 and vice versa. On the other hand, looking at a Leontief production function, the product amount y_0 can only be produced with one combination of inputs, namely the combination (z_{10}, z_{20}), as shown in Fig. 18.1b.

It is, of course, also possible to produce the product amount y^0 with the use of larger amounts of input than that which corresponds to (z_{10}, z_{20}). (If so, the producer can choose to only use the amount (z_{10}, z_{20}) and then discard the surplus of z_1 and z_2). Therefore, the actual production possibility area for production of the amount y^0 is equal to all input combinations "north-east" of (z_{10}, z_{20}), an area that is limited by the L-shaped curve with a corner in (z_{10}, z_{20}). This L-shaped curve is thus formally the isoquant for production of the product amount y^0. However, in practice, only the corner point (z_{10}, z_{20}) is relevant. Other points on the isoquant will entail a larger consumption of either z_{10} or z_{20} without the possibility of reducing the consumption of the other input in question.

The Leontief production technology is described in more detail in Fig. 18.2. The upper part of Fig. (18.2a) illustrates how the production increases linearly for each of the three inputs as a function of the amount of input. This linear increase

S. Rasmussen, *Production Economics*, DOI: 10.1007/978-3-642-14610-7_18,

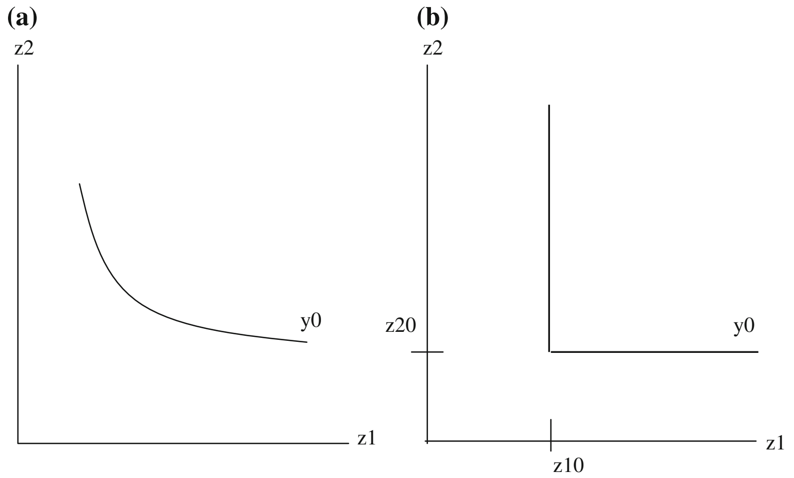


Fig. 18.1 Isoquants in **a** Neoclassical and **b** Leontief production technology

presupposes, however, that the other inputs are available in sufficient quantities, as each unit of the product y requires *both* a_1 units of z_1, and a_2 units of z_2, and a_3 units of z_3. With a restriction of only available b_2 units of z_2, the maximum production is equal to y_{max}, as shown in the figure. Hence, it is the input that is available in the smallest amount, relatively speaking, which effectively limits the production. Hence, the production maximum y_{max} is determined by the input amount b_2, and even though there are surplus units of z_1 and z_3 (r_1 respectively r_3 units) at this production level, it is not possible to produce more than that which corresponds to b_2. This is due to the so-called "minimum law" which is the reason for the use of the minimisation operator in the mathematical formulation of the production function shown under item (c) in Fig. 18.2.

Apart from the lack of a substitution possibility as just described, the Leontief production function is characterised by being linear homogeneous (homogeneous of degree one). This means that if the consumption of the input factors z_2 and z_3 when producing 1 unit of y is a_2 and a_3, then the consumption when producing 2 units of y will be $2a_2$ and $2a_{3,}$ as outlined under item (b) in Fig. 18.2.

The use of the Leontief production function is illustrated by the following example. Presume that the production of one unit of y requires 2 units of z_1 and 1.5 unit of z_2. There are 12 units of z_1 and 15 units of z_2 available. The production is thus equal to:

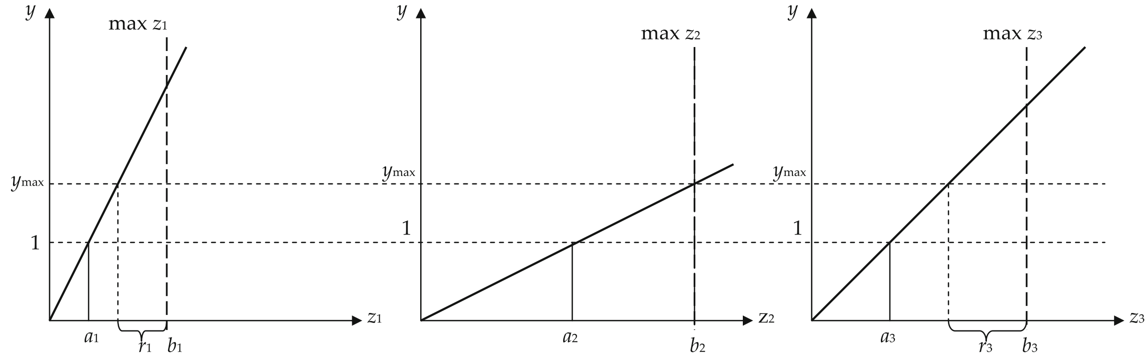
$$y = \min\{(1/2)12, (1/1.5)15\} = 6 \tag{18.1}$$

and it is, thus, the limited amount of input z_1 which effectively restricts production.

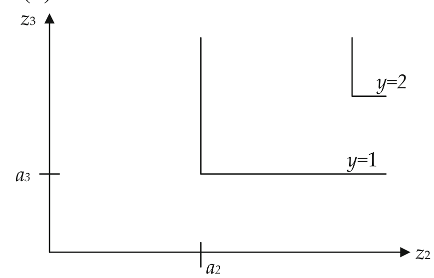
Let us then find out how production will vary with the supplied amount of z_1, when this input is available in any amount, but still with a maximum of 15 units of z_2.

The described relationships can also be illustrated by the well-known production function with one output and only one input. If, for example, it is decided to

(a)



(b)



(c)

$$y = \min\left\{\frac{1}{a_1}z_1, \frac{1}{a_2}z_2, \frac{1}{a_3}z_3\right\}$$

$$= \min\{\alpha_1 z_1, \alpha_2 z_2, \alpha_3 z_3\}$$

in which a_j is input z_j per unit output y

and where α_j is output y per unit of input z_j

Fig. 18.2 A graphical and mathematical representation of Leontief production technology. **a** An example of the production of output y with three inputs z_1, z_2, and z_3. **b** Isoquant with two inputs, z_2 and z_3. **c** Mathematical formulation of Leontief production function

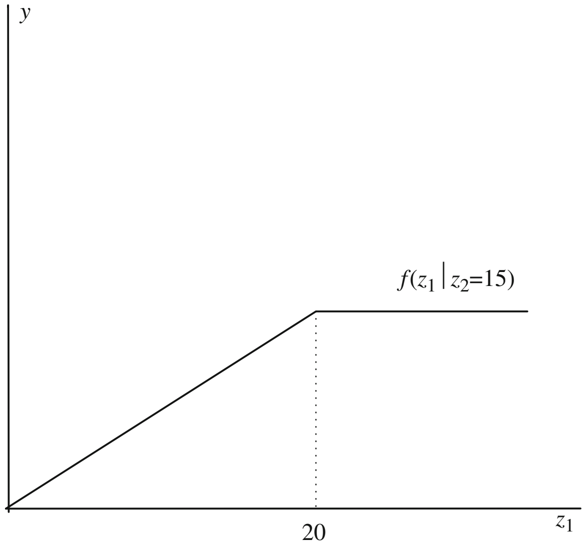


Fig. 18.3 Leontief production technology

keep the amount of input z_2 constant at 15 units, the relationship between y and z_1 can be illustrated as shown in Fig. 18.3.

As can be seen, this is a piecewise linear production function. As long as there are sufficient amounts of z_2 (here 15 units) compared to input 1, it represents a linear increase in production with an increasing supply of z_1. However, at a supply of z_1 more than 20 units, z_2 becomes restricting, and the production function continues horizontally.

Finally, the Leontief production function is linear homogeneous. Doubling the original input amounts (12 units of z_1 and 15 units of z_2) produces $y = \min\{(1/2)24, (1/1.5)30\} = 12$. Hence, a doubling of all inputs results in a doubling of the output.

The use of the Leontief production function can seem to entail considerable restrictions compared to the more flexible neoclassical production model. However, this is not necessarily the case, as demonstrated in the following. Firstly, the use of the Leontief production function makes it possible to use a more powerful optimisation tool, namely Linear Programming. Secondly, the lack of a substitution possibility between the input factors in the base model can easily be remedied by the combination of more Leontief production functions. Finally, it is also possible to outline decreasing as well as increasing scale elasticity by the combination of more Leontief production functions.

18.2 The Production Process

The linear programming model is, in its general form, based on the concept of a production process.

A *production process*, or simply a process, can be defined as a series of numbers, a vector, specifying amounts of input and output which belong together. The example in the introduction (formula (18.1)), which describes a production process with two inputs and one output, can thus be described by the vector:

$$\mathbf{A} = \begin{pmatrix} -1 \\ 1/\alpha_1 \\ 1/\alpha_2 \end{pmatrix} \tag{18.2}$$

in which the number 1 indicates that this is a production process that produces one unit of output by the use of $1/\alpha_1$ units of input z_1, and $1/\alpha_2$ units of input z_2. The minus sign in front of the number 1 indicates that the process *produces* the good in question (here, in a quantity of one unit), while the positive values in the two subsequent elements of the vector indicates that the process *consumes* the goods in question.

An arbitrary production process can, more generally, be written as:

$$\mathbf{A} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix} \tag{18.3}$$

This general formulation can e.g. be describing a production process that produces one or more outputs with the application of one or more inputs. A positive element indicates a consumption of input, while a negative element indicates a production of output. The definition of what constitutes input and output is related to a given process, and what in one process is defined as an output, can be defined as an input in another process.

The elements of the vector (the a's) are referred to as *technical coefficients* and are, hence, expressions of consumption and production when the process is carried out on a unit level.

The process can be carried out with varying *intensity*. In the following, intensity is expressed by a scalar x. The process is said to be carried out with the intensity x when the total consumption of input and the production of output is determined by:

$$\mathbf{A} = \begin{pmatrix} a_1 x \\ a_2 x \\ \vdots \\ a_k x \end{pmatrix} \tag{18.4}$$

Two processes are said to be identical if they produce the same products and use the same input factors in the same ratio. Hence, the processes $\mathbf{A}_1$ and $\mathbf{A}_2$ are identical if it is true that:

$$\mathbf{A}_1 = t\mathbf{A}_2, \tag{18.5}$$

in which t is a constant ($t > 0$).

We will now look at a simple process that produces one output with the use of two inputs:

$$\mathbf{A}' = [a_1, a_2, a_3] \tag{18.6}$$

According to (18.5), all technical coefficients can be multiplied by the same numbers as this can be used to change the process level. If a_3 is set to indicate the production, it is possible to scale the process by multiplication with the numerical value $1/a_3$ so that it corresponds to the production of one unit ($a_3 = -1$). If the letter z is again used to refer to the total consumption of input, the total consumption of input 1 and input 2, respectively, will be given by the following when the process is carried out at the intensity x:

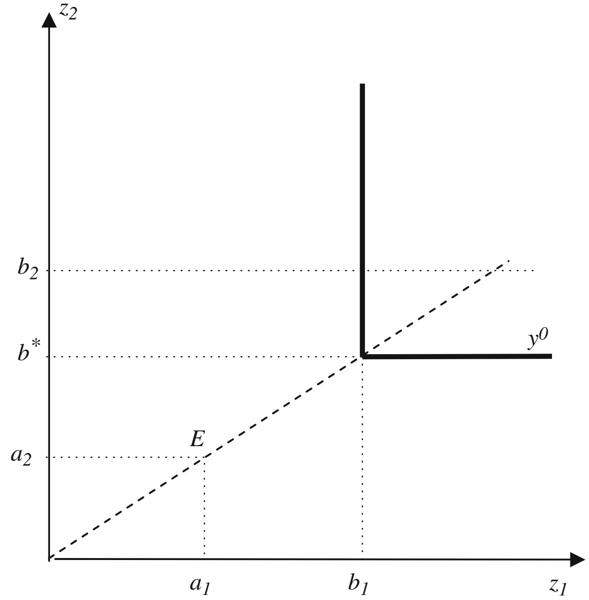
$$z_1 = a_1 x$$
$$z_2 = a_2 x$$

These two equations constitute a parameter representation of a straight line in a z_1–z_2 diagram. The equation of the line is produced by eliminating the parameter x, and the relationship between the total consumption of input 1 and input 2 is:

$$z_2 = (a_2/a_1) z_1$$

The line is drawn in Fig. 18.4 which constitutes a so-called *factor diagram*. Each point on this straight, dotted line through the origin indicates a specific production and a specific consumption of input. The input consumption increases proportionally with the production, and the ratio between the two inputs is always constant. All

Fig. 18.4 Factor diagram for production process



points on the line belong to the same process, while a point outside the line belongs to another process. Point (a_1, a_2) indicates the input consumption at the process unit level, and the distance $0E$ is a measure of the unit production ($y = a_3 = -1$).

As can be seen by comparison with Fig. 18.1b, the process line in Fig. 18.4 is identical to the dotted expansion line of the Leontief production function. A production that is carried out according to a Leontief production function is said to be *linear limitational*. The classical example of a linear limitational production is the production of simple chemical products, such as e.g. hydrochloric acid which is produced by combining one hydrogen with one chlorine atom. The use of one of these two elements in a ratio other than 1:1 will result in a waste. If one or both of these factors are only present in a limited amount, production will be restricted by the factor that gets used up first. Hence, if b_1 and b_2 represent the available amount of the two inputs (see Fig. 18.4), then production will be restricted by input 1, as it is not possible to produce more than that which corresponds to the factor amount b_1. If additional amounts of input 1 are made available, it will be possible to increase production until it becomes restricted by input 2 at the factor amount b_2.

When one or more inputs are available in limited amounts, production will be subject to the restriction that the total consumption of input cannot exceed the available amount. This can be expressed by the inequalities:

$$\begin{aligned} a_1 x &\leq b_1 \\ a_2 x &\leq b_2 \end{aligned} \tag{18.7}$$

These inequalities can be re-written to form equations by the introduction of so-called *slack variables*. If the slack variable for input 1 is referred to as r_1 and the slack variable for input 2 as r_2, (18.7) can be re-written as:

$$\begin{aligned} a_1 x + 1r_1 &= b_1 \\ a_2 x + 1r_2 &= b_2 \end{aligned} \tag{18.8}$$

If Fig. 18.4 is again used for illustration, then the input amounts b_1 and b_2 will result in a production that corresponds to the isoquant y^0, and the slack variable r_1 in (18.8) will be zero, while the slack variable r_2 will be equal to b_2–b^* in Fig. 18.4.

18.3 The Financial Result of the Production Process

As described in the above section, a production process is characterised by a vector of technical coefficients:

$$\mathbf{A} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix} \tag{18.9}$$

The financial result of the process can be calculated if the prices of input and output are known. For this purpose, a *price vector* corresponding to the process is defined as:

$$\mathbf{P} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{pmatrix} \tag{18.10}$$

This price vector should be interpreted in the way that each element of a_{i} in the process vector is related to an element of p_i in the price vector. Hence, the price vector contains the price per unit for input as well as output. As the a's have a negative sign when related to output and a positive sign when related to input, the *financial result of the production process c* will be given by:

$$c = -\mathbf{P}'\mathbf{A} = -p_1 a_1 - p_2 a_2 - \cdots - p_k a_k \tag{18.11}$$

If the process is carried out at the intensity x, *the total financial result* will be:

$$cx = -\mathbf{P}'\mathbf{A}x = -p_1 a_1 x - p_2 a_2 x - \cdots - p_k a_k x \tag{18.12}$$

18.4 Multi-process Models

As mentioned in the introduction, a combination of more Leontief production processes would result in considerable flexibility when illustrating a company's production. Therefore, we will now examine the so-called *multi-process model* in detail.

Presume that several different production processes can be used by a company to produce one or more products. If e.g. there are a total of n possible production processes available, the production model for the company will consist of the following n processes:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & & \mathbf{A}_{\mathrm{n}} \\ a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \end{pmatrix} \tag{18.13}$$

It is now necessary to use two indices whereby the first indicates the number of the input or output in question, while the other indicates the process number. The possibility of a good being produced in one process and consumed in another process is not excluded. Whether a good is an input or an output in the individual process depends, as before, on the sign of the coefficient.

As can be seen, the company's total production possibilities are now indicated by a *matrix* with the columns consisting of the technical coefficients of each individual production process. In the following, *the letter* $\mathbf{A}$ *will be used to refer to this matrix of technical coefficients*. The dimension of this matrix is $(k \times n)$, whereby n is the number of production processes and k is the number of inputs and outputs.

As before, the letter x is used to describe the intensity with which a process is carried out. If x_i is defined as the intensity of the production process i, the company's total intensity can be described by the *intensity vector:*

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

When the company's production is carried out at the intensity $\mathbf{x}$, the total consumption of input and the production of output will be given by the vector:

$$\mathbf{Ax} = \mathbf{A}_1 x_1 + \mathbf{A}_2 x_2 + \cdots + \mathbf{A}_n x_n \tag{18.14}$$

If c_i is used to refer to the financial result of the unit process i (where c_i is defined in a similar way as c in (18.11)), the company's total financial result z can be calculated as:

$$z = \mathbf{c}'\mathbf{x} \tag{18.15}$$

in which $\mathbf{c}'$ is now equal to the vector:

$$\mathbf{c}' = [c_1, c_2, \ldots, c_n] \tag{18.16}$$

18.5 Substitution Between Input Factors

As mentioned in the introduction, the individual Leontief production function (within each production process) does not allow for the substitution between inputs. Inputs are used in given fixed ratios, as illustrated by the corner points in the L-shaped isoquants in Fig. 18.1b. In the neoclassical model, however, substitution is a possibility, as illustrated by the isoquant in Fig. 18.1a, and the final choice of input combination will depend on the ratio between the input prices.

In the real world, you will often encounter productions which are characterised by the input only being used in a given fixed ratio. The production of hydrochloric acid has been mentioned previously. Here the two inputs, hydrogen and chlorine, are always used in the ratio 1:1. In agriculture, there are also productions with fixed ratios between the input factors. Hence, harvesting cereal crops with a combine harvester requires both a combine harvester, and a person to drive the combine

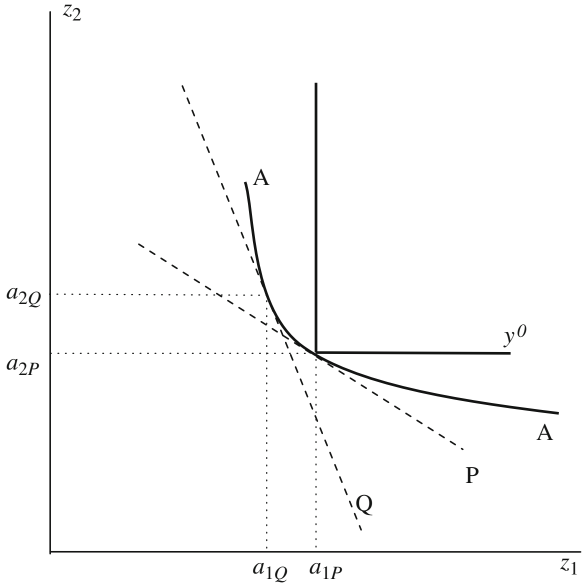


Fig. 18.5 Leontief and neoclassical isoquant

harvester, and these two inputs will always be used in the ratio 1:1. In such cases, the Leontief production function will provide an accurate representation of the actual production conditions. The manufacturing of cars is another example, during which tyres and steering wheels are always used in the ratio 4:1.

In other cases, it will be possible to substitute between inputs. If the Leontief production function is used as the model in such cases anyway, then the model will not correspond to reality.

We will now try to determine whether it is possible to use the Leontief production function as a model in such cases after all.

Presume that a production process is actually of the neoclassical type whereby the production of output y^0 can be carried out by different combinations of the inputs z_1 and z_2, as illustrated by the isoquant A–A in Fig. 18.5. However, instead a Leontief production function is used to illustrate the production conditions represented by the L-shaped isoquant with the corner in the point (a_{1P}, a_{2P}).

If the price ratio between input z_1 and input z_2 (w_1/w_2) corresponds to the absolute value of the slope of the isocost line P, then the input combination (a_{1P}, a_{2P}) will, in reality, be the cost minimising combination of z_1 and z_2 at production of y^0. By using a Leontief model in which z_1 and z_2 are used in combination (a_{1P}, a_{2P}), *no* error is committed as this combination would have been the optimal choice anyway if the more realistic neoclassical model had been used.

However, if the price ratio between the two inputs had been given by the slope of the isocost line *Q*, then using a Leontief production function with the input combination (a_{1P}, a_{2P}) would have resulted in an error as the cost minimising combination in reality would have been (a_{1Q}, a_{2Q}).

How can the occurrence of such problems be avoided when the Leontief production functions are preferred under conditions when there is, in fact, a possibility of substitution, and when the optimal combination of inputs will, therefore, depend on the applicable price ratios?

Solution Suggestion 1 The combination of input factors is determined according to the general neoclassical optimisation principles, i.e. the ratio between the relevant inputs is determined as the input combinations which satisfy:

$$\mathrm{MRS}_{ij} = -w_i/w_j \quad (\text{all } i \text{ and } j) \tag{18.17}$$

whereby MRS is the marginal substitution rate between input i and input j. Hence, the condition states that input i and input j should be used in a combination where the marginal substitution rate is equal to the price ratio. Graphically, this corresponds to the tangent point of the isocost line and the isoquant (see Fig. 18.5).

This solution suggestion presupposes two things.

Firstly, knowledge of the relevant price of the input factors is required. This is naturally no problem when using variable input factors. However, the problem here can be that these prices vary over time. Thus, with each new price change, the ratio between the technical coefficients in the Leontief production process should be established again. When it comes to the fixed input factors, it may be more difficult to establish a price. The price of fixed factors should be determined from an opportunity cost perspective (shadow price). And shadow prices of fixed factors are *not* something that can be found directly in price statistics. We will return to this problem later.

Secondly, the actual, underlying production function is required to be homogeneous of first degree. Real world production functions are not necessarily homogeneous of first degree; however, used as a local approximation, such a precondition can often be justified.

Solution Suggestion 2 This solution is of a more general nature. It is based on an explicit formulation of substitution possibilities by the use of *several* Leontief production functions.

For the sake of simplification, we will look at a production process that produces one output with the use of two inputs. Presume, furthermore, that the product in question can be produced with different combinations of the two inputs. Two of these possibilities are given by the production processes $\mathbf{A}_1$ and $\mathbf{A}_2$ given by:

$$\mathbf{A}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ -1 \end{pmatrix}$$

and

$$\mathbf{A}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ -1 \end{pmatrix}$$

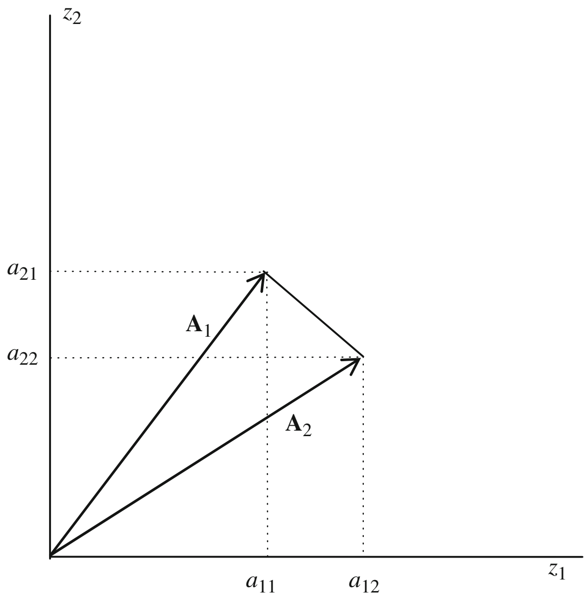


Fig. 18.6 Factor diagram

As can be seen, these processes are scaled so that they both produce an output of one product unit (the specific scaling is not crucial). The two processes are drawn in the same factor diagram in Fig. 18.6.

If the production process $\mathbf{A}_1$ is carried out with the intensity 1 ($x_1 = 1$), the production will be equal to one unit. This is also the case if the process $\mathbf{A}_2$ is carried out with the intensity 1 ($x_2 = 1$).

Now presume that it is also possible to produce the product in question by c*ombining the two processes*. Presume, furthermore, that the processes are *divisible* and *additive*. Divisible is understood to mean that the processes can be carried out with an arbitrary intensity. In reality, this assumption implies that the inputs used are arbitrarily divisible. Additive is understood to mean that the total consumption of input is equal to the sum of the consumption of input for each individual process, and that the total production is equal to the sum of the production of each individual process.

With these assumptions, the total consumption of input and production of output will be equal to:

$$\begin{pmatrix} z_2 \\ z_1 \\ -y^0 \end{pmatrix} = \begin{pmatrix} a_{12} \\ a_{22} \\ -1 \end{pmatrix} x_1 + \begin{pmatrix} a_{11} \\ a_{21} \\ -1 \end{pmatrix} x_2 \tag{18.18}$$

in which x_1 is the intensity for process 1 and x_2 is the intensity for process 2, and where y^0 is the total product amount (here, equal to x_1+x_2).

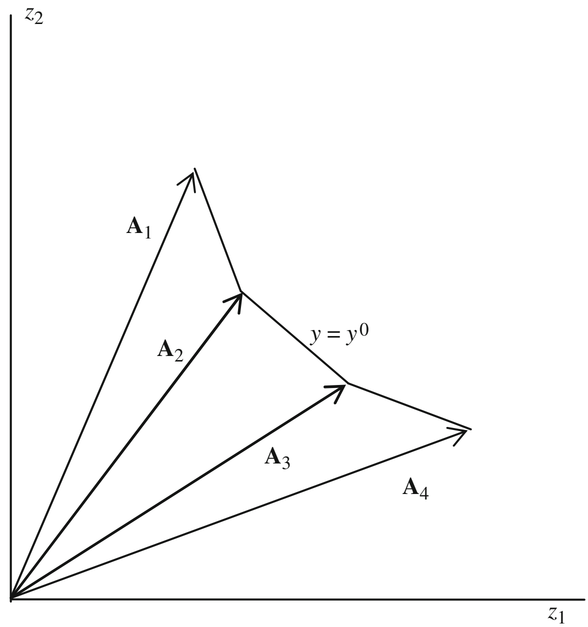


Fig. 18.7 Isoquant illustrated with four processes

If we only consider the value of the intensities x_1 and x_2 defined by:

$$x_1 + x_2 = 1 \quad x_1 \geq 0 \; x_2 \geq 0 \tag{18.19}$$

this is a so-called *convex combination* of the two process vectors $\mathbf{A}_1$ and $\mathbf{A}_2$. If x_1 and x_2 are traced through all permissible values, cf. the conditions in (18.19), it will correspond to the illustration of the straight line from (a_{11}, a_{21}) to (a_{12}, a_{22}) in Fig. 18.6. All points on this line correspond to a product amount of y^0 (here, one unit) and it is, thus, an isoquant corresponding to this production.

The model can be generalised. By combining several processes for the production of one product, an approximated illustration of the neoclassical isoquant can be produced. By combining e.g. 4 processes $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$, and $\mathbf{A}_4$, the isoquant $y = y^0$ can be illustrated, as shown in Fig. 18.7.

Hence, solution 2 provides the possibility of an explicit illustration of substitution possibilities in the linear programming model through a combination of several Leontief production functions.

The choice of one or the other of these solution methods is a matter of preference and will depend on the specific situation. This will be discussed in further detail in Chap. 19.

Chapter 19
Production Planning in the Linear Programming Model: Linear Programming

19.1 Introduction

This chapter discusses how the linear programming model that was introduced in Chap. 18 can be used as a basis for production planning.

Using the concepts already described in Chap. 18, the *production planning problem* can be briefly defined as the act of determining the value of the intensity vector $\boldsymbol{x}$ which is used to maximise the company's total financial result, i.e.

$$\max_{\mathbf{x}}\{\mathbf{c}'\mathbf{x}\} \tag{19.1}$$

where $\mathbf{c}$ is a $[n \times 1]$ vector of the financial results of the unit processes, and $\mathbf{x}$ is a $[n \times 1]$ vector of intensities for the possible production processes.

As can be seen, (19.1) is a linear function with no maximum (the c_i's are presumed to be >0). The production planning problem is, therefore, only relevant if the company's production is subject to certain restrictions.

In the short and medium run, the production will be restricted by the quantity of fixed input factors available to the company. The maximisation in (19.1) must thus be carried out under the constraint that the consumption of the input does not exceed the available quantity.

19.2 Fixed and Variable Inputs

Within production economics, the definition of fixed and variable inputs is related to *the length of the planning period*. Hence, it is possible to talk of planning in the *short run* if one or more of the input factors are fixed, and to talk about planning in the *long run* if all input factors are variable. The question of fixed vs. variable inputs is, in this connection, related to whether it is technically possible to make any changes to the input amount within the considered period.

S. Rasmussen, *Production Economics*, DOI: 10.1007/978-3-642-14610-7_19,

The following will primarily consist of a discussion of planning in the short run. Hence, we will primarily be concerned with problems of planning when one or more inputs are fixed, and when one or more inputs are variable.

The economic version of the "technical" definition of fixed and variable inputs is:

Fixed inputs:	Purchase price = ∞. Sales price = 0.
Variable inputs:	Purchase price = Sales price = Market price.

These definitions specify more explicitly that a company's fixed inputs are those which have no alternative usage outside the company or which, regardless of the sales price, are not for sale. Hence, these fixed inputs can be used by the company without any costs (the price per unit equals zero). The use of variable inputs will, on the other hand, entail costs corresponding to the market price.

19.3 Optimisation with Fixed and Variable Inputs

As described in Chap. 18 (18.3), a production process—e.g. $\mathbf{A}_i$—is defined by the following vector of technical coefficients:

$$\mathbf{A}_i = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \\ a_{m+1,i} \\ \vdots \\ a_{ki} \end{bmatrix} \tag{19.2}$$

in which negative a's indicate the production of output, while positive a's indicate the consumption of input.

Of the total k goods, which are input and which are output depends on the specific situation. Here, the first m good is assumed to be a fixed input, i.e. $a_{1i},\ldots, a_{mi}$ indicate the consumption of fixed inputs per process unit. The remaining k-m elements, i.e. $a_{m+1,i},\ldots, a_{ki}$ are variable inputs and/or output per process unit.

A price vector corresponding to the k goods that are part of the production process can now be defined as follows:

$$\mathbf{P} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ p_{m+1} \\ \vdots \\ p_k \end{bmatrix} \tag{19.3}$$

As the first m inputs are fixed inputs, the prices of these inputs are 0 which can be seen from the above price vector.

The total financial result of the unit process c_i is then equal to:

$$c_i = -\mathbf{P}'\mathbf{A}_i = -p_{m+1}a_{m+1,i} - \cdots - p_k a_{ki} \tag{19.4}$$

The expression in (19.4) is equal to the product value minus the variable costs, and the total financial result of the unit process is, therefore, equal to the *gross margin* of the process.

Similarly, the financial result (the gross margin) of the unit process can be calculated for the company's other processes ($i = 1,\ldots,n$). The company's total financial result (the gross margin) can thus be calculated as:

$$\mathbf{c}'\mathbf{x} = c_1x_1 + c_2x_2 + \cdots + c_nx_n \tag{19.5}$$

However, the fact that a number of the company's inputs are fixed inputs must be taken into consideration. The production is, thus, subject to the restriction that the consumption of the fixed inputs cannot exceed the available amount. Thus, the overall *optimisation problem* becomes equal to:

$$\max\{c_1x_1 + c_2x_2 + \cdots + c_nx_n\}$$

subject to the constraints:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1\mathrm{n}}x_{\mathrm{n}} &\le b_1 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &\le b_m \end{aligned} \tag{19.6}$$

where $[b_1,\ldots, b_m]$' is the vector of the company's fixed inputs. If the vector and matrix concepts defined in Chap. 18 are used, and the concept **b** is introduced for the vector $[b_1,\ldots, b_m]$', the optimisation problem in (19.6) can be briefly described as:

$$\max \boldsymbol{c}'\mathbf{x}$$

subject to the constraints:

$$\mathbf{Ax} \le \mathbf{b} \tag{19.7}$$

in which $\mathbf{c}'\mathbf{x}$ is referred to as *the objective function* (or the *criteria function*), **A** is the matrix of technical coefficients, and the vector **b** is referred to as *the right hand side*.

The problem described can be solved by means of linear programming (LP).

19.4 An Example

Presume that a farmer has the possibility of growing *two crops* (products), e.g. barley and oilseed rape. The production requires *four inputs*, consisting of *two fixed*, input 1 (good 1), and input 2 (good 2) (e.g. land and labour), and *two* *variable*, input 3 (good 3), and input 4 (good 4) (e.g. fertiliser and pesticides). The two products are referred to as good 5 (barley) and good 6 (oilseed rape).

The price of the variable inputs is MU 2 per kg of input 3 (fertiliser) ($p_3 = 2$), and MU 250 per litre of input 4 (pesticides) ($p_4 = 250$). The price of the two products is MU 1.20 per kg for barley ($p_5 = 1.20$), and MU 3.00 per kg for oilseed rape ($p_6 = 3.00$). As good 1 and 2 are fixed inputs, the prices p_1 and p_2 are equal to 0 (zero).

The price vector $\boldsymbol{P}$ is therefore equal to:

$$\mathbf{P} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 2 \\ 250 \\ 1.20 \\ 3.00 \end{pmatrix}$$

The process unit for barley is fixed at one hectare. Hence, the process unit for barley requires 1 ha of land ($a_{11} = 1$). Furthermore, the growing of barley per hectare is assumed to require 12 h of labour ($a_{21} = 12$), 400 kg of fertiliser ($a_{31} = 400$), and 4 l of pesticides ($a_{41} = 4$). Finally, the yield per hectare is equal to 4,000 kg ($a_{51} = -4{,}000$).

The process unit for oilseed rape is also fixed at one hectare. Hence, the process unit for oilseed rape requires 1 hectare of land ($a_{12} = 1$). Furthermore, the growing of oilseed rape per hectare is assumed to require 10 h of labour ($a_{22} = 10$), 500 kg of fertiliser ($a_{32} = 500$), and 6 l of pesticides ($a_{42} = 6$). Finally, the yield per hectare is equal to 1,800 kg ($a_{62} = -1{,}800$).

Hence, the two processes have the following appearance:

$$\mathbf{A}_1 = \begin{pmatrix} 1 \\ 12 \\ 400 \\ 4 \\ -4{,}000 \\ 0 \end{pmatrix} \quad \mathbf{A}_2 = \begin{pmatrix} 1 \\ 10 \\ 500 \\ 6 \\ 0 \\ -1{,}800 \end{pmatrix}$$

The total financial result (gross margin) of the two processes is then equal to:

$$c_1 = -\mathbf{P}'\mathbf{A}_1 = \text{MU } 3{,}000$$

$$c_2 = -\mathbf{P}'\mathbf{A}_2 = \text{MU } 2{,}900$$

Finally, the amount of fixed input is assumed to be equal to 50 ha of land and 550 h of labour. Thus, the restriction vector $\boldsymbol{b}$ is equal to:

$$\mathbf{b}' = [b_1, b_2] = [50, 550]$$

The production planning problem can then be formulated as:

$$\max\{3,000x_1 + 2,900x_2\}$$

subject to the constraints:

$$1x_1 + 1x_2 \leq 50$$

$$12x_1 + 10x_2 \leq 550$$

19.5 Quasi-Fixed Input

The optimisation problem just discussed is formulated under the precondition that there are only two types of input: fixed or variable. Furthermore, it is assumed that the amounts of product produced are sold by the company at the prices given by the price vector **P**.

These preconditions will now be modified. Firstly, we will examine the optimisation problem when there are *quasi-fixed inputs* involved as well as the fixed and variable inputs. Furthermore, we will look at the problem formulation when the products from a production process are not sold but are included as inputs to another production process.

Quasi-fixed inputs can be defined as inputs whereby the decision-maker reacts in accordance with the following price ratios:

$$\text{Quasi-fixed input: } \infty > \text{Purchase price} > \text{Sales price} > 0$$

These relationships should be interpreted in the following way.

Initially, the company has a given amount of the input in question. If the *internal value* (the marginal value by use of the input in question in the company's production (the shadow price)) of the input in question is greater than the purchase price, then the company is ready to buy more units. If the internal value is lower than the sales price, the company is ready to dispose of some of the units initially available. If the internal value, on the other hand, is found somewhere between the sales price and the purchase price, no changes are desirable.

Hence, the following can be said to be true for the quasi-fixed inputs:

$$\text{Internal value} > \text{Purchase price} \Rightarrow \text{Variable input}$$

$$\text{Internal value} < \text{Sales price} \Rightarrow \text{Variable input}$$

$$\text{Purchase price} > \text{Internal value} > \text{Sales price} \Rightarrow \text{Fixed input}$$

Quasi-fixed inputs are therefore also referred to as *conditional fixed inputs or conditional variable inputs* as the input in question is fixed or variable as a *condition* of the internal value of the input in question.

Initially, the internal value of the input in question is unknown (only when the optimisation problem in (19.7) has been solved, will the shadow prices of the fixed factors be known). When formulating the optimisation problem, you will not know beforehand whether to treat the input in question as a fixed input, as a variable input with a price equal to the purchase price, or as a variable input with a price equal to the sales price.

To allow for the possibility of all three situations, the production model can be expanded to include processes that have the sole purpose of buying or selling the input in question.

Presume that input m, which was a fixed input in the previous section, is now a *quasi-fixed input*. To allow for the possibility of buying more units or selling some of the units, two more processes are now formulated, $\boldsymbol{A}_{n+1}$ and $\boldsymbol{A}_{n+2}$, respectively, as follows:

$$\mathbf{A}_{n+1} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad \mathbf{A}_{n+2} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{19.8}$$

The process $\boldsymbol{A}_{n+1}$ is a process that *buys* one unit of input m at unit level. Hence, this process *results in* one unit of good m for the company. (Hence, the coefficient -1 in row m). The process $\boldsymbol{A}_{n+2}$ is a process that *sells* one unit of input m at unit level. Hence, this process *consumes* one unit of good m (Hence, the coefficient +1 in row m).

The purchase price when buying extra units of input m is p_{mk}, and the sales price when selling the available units of input m is p_{ms}. The financial result c_{n+1} for the process $\boldsymbol{A}_{n+1}$ is, therefore, equal to $-p_{mk}$, and the financial result c_{n+2} for the process $\boldsymbol{A}_{n+2}$ is equal to p_{ms}.

The overall optimisation problem can then be formulated as:

$$\max\{c_1x_1 + c_2x_2 + \cdots + c_nx_n - p_{mk}x_{\mathrm{n+1}} + p_{ms}x_{\mathrm{n+2}}\}$$

under the constraints:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &\le b_1 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n - 1x_{n+1} + 1x_{n+2} &\le b_m \end{aligned} \tag{19.9}$$

whereby $x_{\mathrm{n+1}}$ and $x_{\mathrm{n+2}}$ are now equal to the intensity of the purchase and sale, respectively, of input m.

When formulating purchase and sales processes for input as shown here for input m, it is possible to illustrate quasi-fixed inputs. The formulation and use of purchase and sales processes will also be useful in other connections, illustrated in the examples in Chap. 20. See also Hazell and Norton (1986).

19.6 Intermediate Goods

Intermediate goods are defined here as goods which are produced as products in one (or more) production process(es), and are used as inputs in another (or several) production process(es). An example of an intermediate good in the operation of a farm is grass, which is a product (output) in the grass production process, and an input in the milk production process.

The usual references are used in the following, and it is now presumed that good k is an output in process number i, and an input in process number j. If the reference from Sect. 19.3 is used, then a_{ki} is negative and the corresponding coefficient in process j (a_{kj}) is positive. It is, furthermore, assumed that the good in question *cannot* be purchased or sold, and that the consumption of this good as an input in process number j, thus presupposes a corresponding production in process number i. (The assumption that it is not possible to purchase or sell the good in question is not crucial. It is always possible to subsequently expand the model to include purchase and sales processes for the good in question in the same way as described in Sect. 19.5 above).

The production processes $\boldsymbol{A}_i$ and $\boldsymbol{A}_j$ have the following appearance:

$$\mathbf{A}_i = \begin{pmatrix} a_{1i} \\ \vdots \\ a_{mi} \\ a_{m+1,i} \\ \vdots \\ a_{k-1,i} \\ -a_{ki} \end{pmatrix} \quad \mathbf{A}_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \\ a_{m+1,j} \\ \vdots \\ a_{k-1,j} \\ +a_{kj} \end{pmatrix} \tag{19.10}$$

as shown, good k is an output in process i ($-a_{ki} < 0$), and an input in process j ($+a_{kj} > 0$).

The original price vector in (19.3) is modified correspondingly:

$$\mathbf{P} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ p_{m+1} \\ \vdots \\ p_{k-1} \\ 0 \end{pmatrix} \tag{19.11}$$

it should be noted that good k is given the price 0 (zero) to economically illustrate that the good cannot be sold (market price = 0).

As in formula (19.5) in Sect. 19.3, the financial result of the processes is calculated as:

$$c_i = -\mathbf{P}'\mathbf{A}_i = -p_{m+1}a_{m+1,i} - \cdots - p_{k-1}a_{k-1,i} \tag{19.12}$$

and:

$$c_j = -\mathbf{P}'\mathbf{A}_j = -p_{m+1}a_{m+1,\mathrm{j}} - \cdots - p_{k-1}a_{k-1,j} \tag{19.13}$$

As can be seen, good k is not included in the calculation of the financial result of the two unit processes at all, not as the product value in process i, or as a variable cost in process j. *Hence, intermediate goods should not be included in the calculation of the financial result of the processes.*

On the other hand, the previously mentioned condition, i.e. that the consumption of good k as an input in the process j must not exceed the production of good k in process i, should be taken into consideration.

This condition can be formulated as:

$$a_{ki}x_i \geq a_{kj}x_j \tag{19.14}$$

The final optimisation model can then be formulated as:

$$\max\left\{c_1x_1 + \cdots + c_ix_i + \cdots + c_jx_j + \cdots + c_nx_n\right\}$$

under the constraints:

$$\begin{aligned} &a_{11}x_1 + \cdots + a_{1i}x_i + \cdots + a_{1j}x_j + \cdots + a_{1n}x_n \leq b_1 \\ &\qquad\qquad\qquad\qquad \vdots \\ &a_{m1}x_1 + \cdots + a_{mi}x_i + a_{mj}x_j + \cdots + a_{mn}x_n \leq b_m \\ &-a_{ki}x_i + a_{kj}x_j \leq 0 \end{aligned} \tag{19.15}$$

whereby the last restriction in the constraints is produced by multiplying (19.14) by -1 and moving $a_{kj}x_j$ over to the left hand side of the inequality.

Hence, the turnover of intermediate goods in the linear programming model is generated by adding a restriction that “keeps track” of this turnover.

The last restriction in the model (19.15) can also be interpreted along the same lines as the other restrictions, i.e. as a condition which makes sure that the (net) consumption of fixed input cannot exceed the available amount. The zero on the right hand side of the inequality thus indicates that the amount of the “fixed” input is initially zero, which is precisely the right interpretation.

With this last interpretation, it is relatively easy to further expand the model to also include the possibility of buying and/or selling the good in question. This is done using the principles described in Sect. 19.5.

19.7 Determination of the Production Model Parameters

Solving the production planning problem:

$$\max \ \mathbf{c}'\mathbf{x}$$

under the constraints:

$$\mathbf{A}\mathbf{x} \leq \mathbf{b} \tag{19.16}$$

presupposes the determination of the three sets of parameters, the vector of gross margins **c**, the matrix of technical coefficients **A**, and the right hand side vector **b**.

The right hand side vector does not give rise to any major fundamental problems. This is a vector of the company's fixed inputs. There might, of course, be practical problems with regard to the choice of the length of the planning period and, thus, with regard to the identification of the inputs that are considered fixed and should, therefore, be included in the right hand side, and with regard to which of the inputs are variable and, thus, should be included in the calculation of the gross margins (the c's).

The vector of gross margins (**c**) includes the financial result of the individual unit processes as elements. As shown by formula (19.4) in Sect. 19.3, the financial result is calculated as the (expected) gross margin of the unit process. This calculation includes the expected product prices, the prices of variable inputs [the p's in (19.3)], technical coefficients in the form of the consumption of variable input factors per process unit, and the product output per process unit [the a's in (19.3)].

There are no obvious fundamental problems related to the determination of the prices of variable inputs and expected product prices, especially when the company is presumed to be in a situation of perfect competition and, thus, is a price taker in both the product, as well as the factor market. There is, of course, a problem of uncertainty involved, as it can be difficult when planning to predict the future price for the product. This uncertainty is a characteristic of agriculture, forestry, plant nursery and other industries, but is not specifically related to the linear programming model, which is why it is not discussed here.

However, the determination of the technical coefficients should be analysed in more detail. These coefficients (the a's) are included both in the calculation of the gross margins and also as elements in the matrix $\boldsymbol{A}$. The fundamental problems associated with the determination of these coefficients are the same in both situations, and they are therefore discussed as one here.

The problem regarding the determination of the technical coefficients—or more precisely: *the ratio between the technical coefficients*—has been briefly discussed in Chap. 18, Sect. 18.5 above. The problem is related to the use of a Leontief production function under conditions where the actual production technology is neoclassical.

The point of reference is a neoclassical production function:

$$y = f(z_1, \ldots, z_k) \tag{19.17}$$

It has been shown above (Chap. 4) that the optimal combination of input is found along the expansion path where:

$$p_i/p_j = \text{MPP}_i/\text{MPP}_j, \tag{19.18}$$

The condition (19.18) is operational in connection with the establishment of the relationship between the unit process' consumption of two variable inputs. In connection with variable inputs, the market prices are of course known—i.e. the p's. Hence, the ratio between inputs i and j should be determined so that the ratio between the marginal products (MPP) is equal to the ratio between the market prices.

If, on the other hand, this concerns the ratio between two fixed inputs, or the ratio between a fixed and a variable input, the problem is more significant. In this case, there is no precise price for fixed inputs. Indeed, there is no relevant price at all at the time of planning. The theoretically correct price to use in (19.18), with regard to fixed inputs, is the shadow price (the internal value). And this value is not known until the linear programming problem has been solved!

This is of course inappropriate. In principle, this problem can be addressed in two ways.

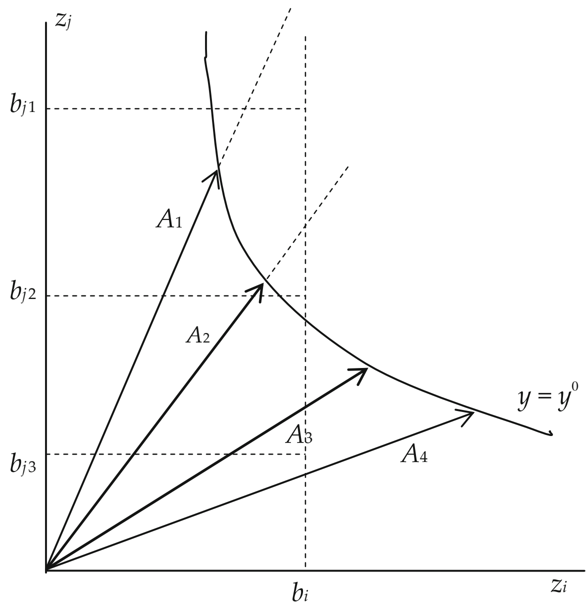
One way is to start by "guesstimating" (read: guess) a shadow price and, on the basis of this, determine the ratio between the technical coefficients for all production processes by use of (19.18). The optimisation problem is then solved by means of linear programming, and the shadow prices for the fixed inputs can be read. If these shadow prices deviate from the original guess, the prices are modified, and (19.18) is again used to modify the calculated ratio between the technical coefficients. The optimisation problem is then solved again. This iterative procedure is continued until a reasonable agreement has been achieved between the prices used in (19.18) and the actual shadow prices.

The other method consists of formulating two or more production processes for each product. The formulation of these processes is done to achieve a reasonably approximated representation of the actual isoquant, as illustrated in Figs. 18.5 and 18.6 in Chap. 18. Thus, the model itself contains the possibility of choosing the process (or combination of processes) that is optimal with regard to the scarcity of the fixed inputs (and, hence, the shadow price).

An attempt to illustrate this procedure has been made in Fig. 19.1. This concerns the ratio between two fixed inputs, z_i and z_j, that are available in the amounts b_i and b_j. It, furthermore, concerns the production of a product that can be produced by use of z_i and z_j in different combinations. According to this, 4 different processes, $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, $\boldsymbol{A}_3$, and $\boldsymbol{A}_4$, which use the two inputs in different ratios, have been formulated. The isoquant for the production of a base amount (the production per process unit) y^0 is also outlined in the figure.

If input z_j is available in a comparatively sufficient amount compared to z_i ($z_j = b_{j1}$ and $z_i = b_i$), production will be optimal at a combination of process $\mathbf{A}_1$

Fig. 19.1 Combination of processes



and $\mathbf{A}_2$. In case of smaller amounts of z_j compared to z_i (e.g. $z_j = b_{j2}$), the optimal process combination will change to $\boldsymbol{A}_2$ and $\boldsymbol{A}_3$. Finally, an even smaller amount of z_j compared to z_i (e.g. $z_j = b_{j3}$) would mean that the optimal process combination would change to $\boldsymbol{A}_3$ and $\boldsymbol{A}_4$.

Hence, this shows that by formulating several processes for the production of a given product, the model will itself choose the process combination which is optimal under the given input ratios. Indirectly, this means that the model chooses the input combination according to the actual shadow prices of the fixed inputs.

The problem will not be discussed in further detail here. The essential point is to use a critical approach to the determination of the technical coefficients of processes. And in the above, instructions have been given as to the key considerations in this connection, and the possible solutions available.

19.8 Interpretation of the Dual Problem

This chapter is concluded with an economic interpretation of the dual formulation and the optimisation of the linear programming model.

Each primal formulation of a linear programming problem (LP problem) has a corresponding so-called dual formulation. The information produced by the use of a simplex algorithm to solve the primal problem precisely corresponds to the information achieved by solving the dual problem. Therefore, there is a ‘free choice' option between formulating and solving the primal or the dual problem (see Hillier and Lieberman 1998).

The same is, of course, true for the linear programming model. The general primal formulation has already been shown in (19.4) and is repeated here:

Primal Formulation:

$$\max\{c_1x_1 + c_2x_2 + \cdots + c_nx_n\}$$

under the constraints:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1\text{n}}x_\text{n} &\le b_1 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &\le b_m \end{aligned} \tag{19.19}$$

The related dual formulation is given by the following:

Dual Formulation:

$$\min\{b_1y_1 + b_2y_2 + \cdots + b_my_m\}$$

under the constraints:

$$\begin{aligned} a_{11}y_1 + a_{21}y_2 + \cdots + a_{m1}y_m &\ge c_1 \\ &\vdots \\ a_{1n}y_1 + a_{2n}y_2 + \cdots + a_{mn}y_m &\ge c_n \end{aligned} \tag{19.20}$$

in which the y's (the dual variables) are equal to the shadow prices (internal values, or marginal values) of the fixed inputs in the primal problem.

With regard to production functions that are homogeneous of degree one, the following is true according to Euler's Theorem (see e.g. Debertin 1986, p. 162–163):

$$z_1\text{MPP}_1 + \cdots + z_m\text{MPP}_m = y \tag{19.21}$$

i.e. that the total production y can be calculated as the product sum of the resource consumption (the z's) and the marginal products (the MPP's).

If both sides of the equal sign in (19.21) are multiplied by the product price p_y, (19.21) can be written as:

$$z_1\text{VMP}_1 + \cdots + z_m\text{VMP}_m = p_y y \tag{19.22}$$

i.e. that the total financial result of the production (the product value on the right hand side) can be calculated as the product sum of the resource consumption and the marginal product values of the resources [the VMP's indicate the marginal product values calculated as the marginal product (MPP) multiplied by the product price (p_y)].

As the linear programming model is homogeneous of degree one, the total financial result can also be calculated as the product sum of the factor consumption and the marginal product values of the factors. However, this is precisely what is indicated in the object function of the dual problem in (19.20). This concludes the description of the interpretation of the well-known LP based result, in that the value of the object function when solving the dual problem is the same as when solving the primal problem, and that the dual variables are in fact equal to the marginal value of the resources (shadow prices).

Finally, the dual problem is described/interpreted verbally. Firstly, as a basis for comparison, the primal problem is described:

The Primal Production Planning Problem:

> Determine the production composition that maximises the company's total financial result when, at the same time, the consumption of fixed inputs is restricted from exceeding the available quantities of fixed resources.

The Dual Production Planning Problem:

> Determine the prices at which the company's fixed resources can be leased out when they are to be leased out at the lowest possible price but, at the same time, at a price that corresponds to the income that could have been achieved if the resources had been used for production within the company.

Hence, as can be seen, the dual production planning problem can be attributed a completely independent economic interpretation, i.e. the interpretation that is related to leasing out the fixed resources instead of using them yourself.

References

Debertin, D. L. (1986). *Agricultural Production Economics*. New York: Macmillan.

Hillier, F. S., & Lieberman, G. J. (1998). *Introduction to Operations Research*. Oakland: Holden-Day, Inc.

Chapter 20
Use of Linear Programming in Practical Production Planning

This chapter discusses how the linear programming model can be used as the basis for a company's operational planning, in practice. The presentation and construction of the LP models will be accounted for through the use of simple examples, primarily from agriculture, and the discussion is aimed at finding a combination of the presented techniques and methods that can then be used to outline a model for the total planning of an agricultural operation. Although the examples are from agriculture, the modelling technique is applicable in other industries.

Section (20.1) contains a presentation of the basic methods for illustrating planning/optimisation problems in a static LP model. This section presupposes a task that consists of finding the optimal production plan for the coming operating year, as the aim is the maximisation of the total gross margin. Hence, the length of the planning period is 1 year, and it is thus the value of the expected planning parameters for this period that will be part of the model. Whether this is a year, a month, or a week is as such not crucial. The essential point is that time is not directly part of the model formulation. However, due consideration should be attributed to all of the relevant planning scope, as many of the actions during the first period (year, month, week) will have implications for the subsequent periods. This is of course especially true, if considerations concerning investment in a fixed asset should be considered when constructing the planning model.

Section (20.2) shows how planning problems spanning over several time periods can be formulated within the framework of an LP model. Such multi-period (or dynamic) models are often necessary in connection with planning within agriculture, when production started in one period is often first finished in subsequent or later periods. Just think of growing grass seeds which entails the sowing of grass in a previous year. Or think of growing forests where the decisions made stretch out many years into the future. The same is true for investment in buildings and machinery. While such dynamic conditions often, with some creativity, can be formulated within the framework of a statistical model, as described in Sect. 20.1, it is often easier and clearer to illustrate dynamic relationships directly in a multi-period model.

S. Rasmussen, *Production Economics*, DOI: 10.1007/978-3-642-14610-7_20,

Section (20.3) deals with integer programming. Integer programming is a technique that can be used to illustrate either–or problems or both–and problems. The technique consists of restricting a number of the LP model's processes so that they only accept integer solutions, and possibly to only accept the values zero (0) or one (1). In this way, it becomes possible, within the framework of an LP model, to illustrate details and restrictions which would otherwise not be possible with continuous processes. The solution to models where certain processes are required to consist of integer variables requires, however, special solution algorithms. The software LINDO includes facilities for handling integer problems.

20.1 Static Models

20.1.1 Introduction

Practical planning problems will only rarely be formulated mathematically. Instead, planning parameters are included in a table which here is referred to as an LP matrix (or an LP tableau) and is constructed similarly to a simplex table. As it is, this reference is in fact used very often, even though the LP matrix lacks a couple of the rows and columns that are found in the simplex table.

In an LP matrix, the model's gross margins (the c_j's) are normally illustrated in the top row of the matrix. The right hand side of the model (the b_i's) is inserted into the column to the right in the matrix. And, finally, the technical coefficients (the a_{ij}'s) are illustrated in the relevant rows and columns.

The restriction system relationships can either be formulated as equations or as inequalities. Inequalities can either be of the type greater than or equal to, or less than or equal to. The latter type is the common type in connection with operations planning problems, however both types are seen. This does not cause any problems with regard to the solution, as most computer software for the solution of LP problems can handle various types of inequalities.

As mentioned in Chap. 19, the elements in the right hand side express the restrictions that must be respected in connection with planning (capacity restrictions, etc.). By using alternative right hand sides, a planning problem can be examined in connection with various requirements concerning the size of the fixed pant (e.g. building capacity, machinery, labour etc.).

The interpretation of three sets of parameters (the a's, the b's, and the c's) has already been discussed in Chap. 19. The interpretation of the mentioned parameters can however vary, depending on the specific model formulation. The criterion function parameters (the c's) normally indicate the financial result (gross margin) of the individual processes per process unit. A positive c_j coefficient represents the gross margin per unit of process no. j, while a negative c_k indicates the variable costs per unit of process no. k. The technical coefficients can be defined as: A positive a_{ij} coefficient indicates the consumption (input) of resource no. i per unit of process no. j. A negative a_{hj} coefficient represents the output (product amount) of product no. h per unit of process no. j. If the coefficient a_{kj} is equal to zero, the process no. j will neither consume nor produce resource or output no. k. Finally, the b's indicate, as mentioned, the constraints which must be considered in connection with planning.

It is evident that the quality of the planning completely depends on how accurately these three sets of parameters are determined, or to put it in another way; on whether the realised parameter values turn out to be equal to the expected values, or whether there will be larger or smaller discrepancies. The possibilities of acquiring sufficiently accurate data for use in planning are, however, not discussed in further detail here.

20.1.2 From Problem to LP Model: an Example

The following relatively simple example comes from Panell (1997).[1]

A farmer has 10 ha available for growing potatoes and/or tomatoes in the combination that yields the highest profit. A special contract with a tomato ketchup factory regarding the supply of tomatoes requires production of at least 2 ha of tomatoes for the factory. After having carried out other tasks, the farmer has the possibility of using 12 h per week for the cultivation of the 10 ha. Each hectare of potatoes requires 2 h per week, while the tomatoes require 0.5 h per week and hectare. Potatoes provide total revenue of MU 4,000 ha^{-1}, while tomatoes provide MU 3,000 ha^{-1}. Fertiliser costs MU 1,000 tonne^{-1} and should be applied in the following amounts: 1 tonne ha^{-1} for potatoes, and 0.5 tonne ha^{-1} for tomatoes.

20.1.2.1 Process Units

First, a process unit should be chosen. There are two processes, growing potatoes and growing tomatoes. As all data are indicated in units per hectare, it seems expedient to choose hectare as the process unit for both processes. However, other units could have been chosen, e.g. units of yield (kg or tonne).

20.1.2.2 Object Function

The profit (the gross margin) is calculated for each of the crops by deducting costs for fertiliser from the total revenue. For potatoes, the total revenue per hectare is MU 4,000 and the cost of fertiliser is MU 1,000 tonne^{-1}. Hence, the gross margin is MU 3,000 ha^{-1}. For tomatoes the total revenue is MU 3,000, and the cost of fertiliser is 0.5 multiplied by MU 1,000. Hence, the gross margin is MU 2,500 ha^{-1}.

---

[1] Panell (1997) contains a lot of good examples and graphical illustrations of LP solutions.

If the number of hectares of potatoes is referred to as K, and the number of hectares of tomatoes as T, the following object function is generated:

$$\begin{aligned}&\text{Max GM}\\&\text{where GM} = 3,000K + 2,500T\end{aligned} \tag{20.1}$$

20.1.2.3 Land

Each process unit of potatoes uses 1 ha of land. This is the same for a process unit of tomatoes. No more than 10 ha of land may be used. The restriction can be formulated as:

$$1K + 1T \leq 10 \tag{20.2}$$

20.1.2.4 Contract

Each process unit of tomato contributes with 1 unit (hectare) for the fulfilment of the contract of 2 ha. The process for potatoes, on the other hand, does not contribute, and the coefficient is therefore 0 (zero). The requirement to fulfil the contract of 2 ha of tomatoes can therefore be written as:

$$0K + 1T \geq 2 \tag{20.3}$$

20.1.2.5 Labour

Each process unit of potatoes uses 2 h of labour, and each process unit of tomatoes uses 0.5 h of labour. The requirement that no more than 12 h of labour must be used per week can therefore be formulated as:

$$2K + 0.5T \leq 12 \tag{20.4}$$

20.1.2.6 LP Matrix

All the conditions mentioned in the problem formulation are now specified in a "language" that can be used in an LP matrix. This is shown in Table 20.1.

Table 20.1 LP matrix for simple LP problem

	Potatoes	Tomatoes	Restriction	Right hand side
Object function (MU)	3,000	2,500	Max	–
Land (hectare)	1	1	≤	10
Contract (hectare)		1	≥	2
Labour (t)	2	0.5	≤	12

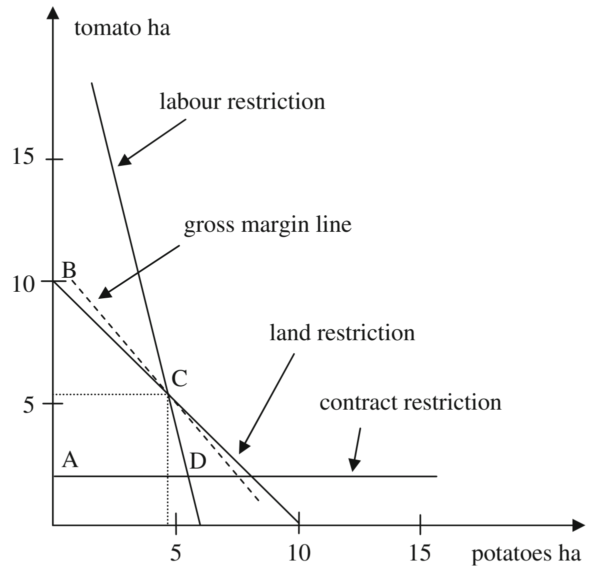


Fig. 20.1 Graphical solution to LP problem

Graphical Solution

The problem can be solved graphically, as shown in Fig. 20.1. The unknown number of process units (number of hectares of potatoes and tomatoes) is plotted along the two axes of the coordinate system. Then the three restriction equations for land, contract and labour, respectively, are drawn in the coordinate system. The restriction lines are produced by Eqs. (20.2), (20.3), and (20.4) where the inequality signs are replaced by equal signs. To fulfil the restriction in (20.2), (20.3), and (20.4), an optimal solution must be found "south-west" of the restriction line for land, "north" of the restriction lines for contract, and "south-west" of the restriction line for labour.

The optimal cultivation plan is found when the gross margin is maximised. The gross margin is illustrated by Eq. (20.1). Solving (20.1) for T as a function of K produces:

$$T = \mathrm{GM}/2,500 - (3,000/2,500)K \tag{20.5}$$

which is the equation for the gross margin line.

Hence, the maximum gross margin is achieved by shifting this line with the slope −3,000/2,500 as far towards the "north-east" as possible, without the line leaving the restriction area delineated be the figure ABCD in Fig. 20.1. As can be seen in the figure, the gross margin line leaves the restriction area exactly at point C. This point C is, thus, the optimal solution (the cultivation plan that provides the largest gross margin). Using the figure, it can be seen that it is optimal to cultivate just less than 5 ha of potatoes and a little over 5 ha of tomatoes.

The solution can also, with a little help from the figure, be determined mathematically. As can be seen from the figure, the optimal solution is found exactly where the land restriction line and the labour restriction line intersect. These two lines have the following equations: $K + T = 10$ and $2K + 0.5T = 12$, respectively. Solving these two equations with the two unknowns produces the solution $K = 4^2/_3$ and $T = 5^1/_3$. Hence, the optimal plan consists of growing $4^2/_3$ ha of potatoes and $5^1/_3$ ha of tomatoes. The total gross margin is $= 4^2/_3 \times 3{,}000 + 5^1/_3 \times 2{,}500 =$ MU 27,333.

A special advantage of linear programming is that it is relatively simple to carry out *sensitivity analyses*. To examine how much the relationship between the gross margins for the two crops can change before the optimal solution changes, the figure illustrates that the slope of the dotted iso-margin line can be change somewhat, without changing the optimal solution (point C). Graphically, this can be illustrated by the dotted line being "tipped" on the point C, without this point changing as the optimal solution. If the line tips *counter clockwise* it will eventually be parallel to the land restriction line. A further counter clockwise swing will imply that point B now becomes the optimal solution. If the gross margin line tips *clockwise* instead, it will eventually be parallel to the labour restriction line. A further clockwise swing will imply that point D now becomes the optimal solution.

To tip the gross margin line clockwise and counter clockwise corresponds to changing the relationship between the gross margins of the two crops. If the gross margin for potatoes is referred to as $GM_{K,}$ and the gross margin for tomatoes as GM_T, the outer limits of how much the gross margin line can be tipped before the optimal solution is changed can be formulated as:

$$-1 \geq -\frac{GM_K}{GM_T} \geq -4 \tag{20.6}$$

in which -1 and -4 are the slopes of the land restriction line and the labour restriction line, respectively.

By first examining how much the gross margin for potatoes (GM_K) can vary, everything else being equal, without changing the optimal solution (C), (20.6) can be formulated as:

$$-1 \geq -\frac{GM_K}{2{,}500} \geq -4 \tag{20.7}$$

Solving (20.7) with regard to GM_K produces:

$$2{,}500 \leq GM_K \leq 10{,}000 \tag{20.8}$$

By instead examining how much the gross margin for tomatoes (GM_K) can vary, everything else being equal, without changing the optimal solution (C), (20.9) can be formulated as:

$$-1 \geq -\frac{3{,}000}{GM_T} \geq -4 \tag{20.9}$$

Solving (20.9) with regard to GM_T produces:

$$750 \leq \mathrm{GM_T} \leq 3,000 \qquad (20.10)$$

Hence, the total gross margin for potatoes can vary between MU 2,500 and 10,000 ha^{-1} according to (20.8) without changing the optimal solution (point C in Fig. 20.1). Examining the gross margin for tomatoes instead, the gross margin can vary between MU 750 and 3,000 ha^{-1} according to (20.10) without changing the optimal solution.

The graphical solution to the LP problem in Fig. 20.1 can also provide the basis for an illustration of the determination of *the shadow prices* of the fixed factors. The shadow price of a fixed factor consists of the extra profit achieved if there had been one more unit available of the input in question.

The present example contains two fixed factors: Land and labour. As can be seen from Fig. 20.1, these two fixed factors constitute a real restriction on the production, as the optimal solution is precisely found at the restriction lines for the two factors (The contract restriction, on the other hand, constitutes no real restriction).

The shadow price for land which, if you remember, is defined as the extra profit achieved by one more hectare of land, can be calculated by changing the land restriction from the present 10 ha to 11 ha, and then solving the LP problem again.

Figure 20.2 contains a new land restriction of 11 ha, compared to Fig. 20.1. As can be seen, the optimal solution is now changed to point C′ where fewer hectares of potatoes, and more hectares of tomatoes are cultivated. The exact solution can be determined mathematically. As before, the optimal solution is found where the land restriction line (new land restriction) and the labour restriction line intersect (point C′). These two lines have the following equations: $K + T = 11$ and $2K + 0.5T = 12$, respectively. Solving these two equations with the two variables T and K, produces the solution $K = 4^1/_3$ and $T = 6^2/_3$. Hence, the optimal plan now consists of growing $4^1/_3$ ha of potatoes and $6^2/_3$ ha of tomatoes. The total gross margin is $= 4^1/_3 \times 3{,}000 + 6^2/_3 \times 2{,}500 =$ MU 29,666.

The gross margin of 10 ha was, as shown above, MU 27,333. If there had been 1 ha more available (11 ha), a gross margin of MU 29,666 could have been achieved. Hence, the extra profit for one hectare is MU 29,666 − 27,333 = 2,333. *The shadow price for land is thus MU 2,333 ha^{-1}*. This amount can be interpreted as the maximum rent that one would be willing to pay if offered to rent one hectare of land more than the 10 ha that are already available.

How many hectares does this shadow price then apply to? As can been directly seen from Fig. 20.2, a continued shifting of the land restriction line towards the "north-east" will result in the optimal solution point C being shifted upwards along the labour restriction line. As long as this change remains linear, each extra hectare will generate an extra profit of MU 2,333. However, at some point the situation will change. When the acreage has increased so much that point C has moved all the way up to the end of the labour restriction line [to the position where the labour restriction line intersects the ordinate (not shown in the figure)], additional acreage will *not* generate any extra profit. In this situation, the production is effectively restricted by the labour restriction, and more land will not improve the earnings as there is no more labour available.

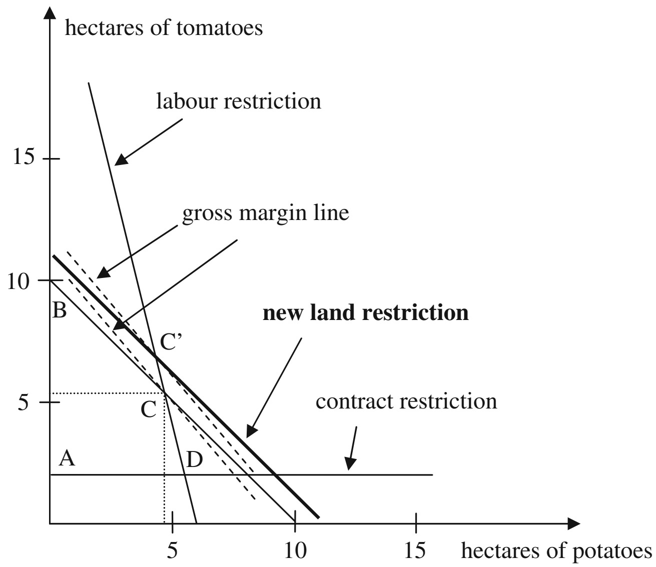


Fig. 20.2 Determination of the shadow price of land

This situation arises when the acreage is expanded to 24 ha of land which are all used for growing tomatoes. Hence, the shadow price of MU 2,333 applies in the interval from 10 to 24 ha, i.e. for 14 (extra) units. For more units than this, the shadow price is equal to zero.

The shadow price can also be calculated by examining how much the profit decreases if the acreage is reduced by 1 unit. As can be seen, a reduction of the acreage by 1 ha from 10 to 9 ha will reduce the profit by MU 2,333. As can be seen from Fig. 20.2, this will be the case until the number of hectares is reduced to 6 ha, which are all used for growing potatoes. A further reduction will imply that the profit decreases by MU 3,000 ha^{-1}, as each hectare that is abandoned generates a gross margin of MU 3,000.

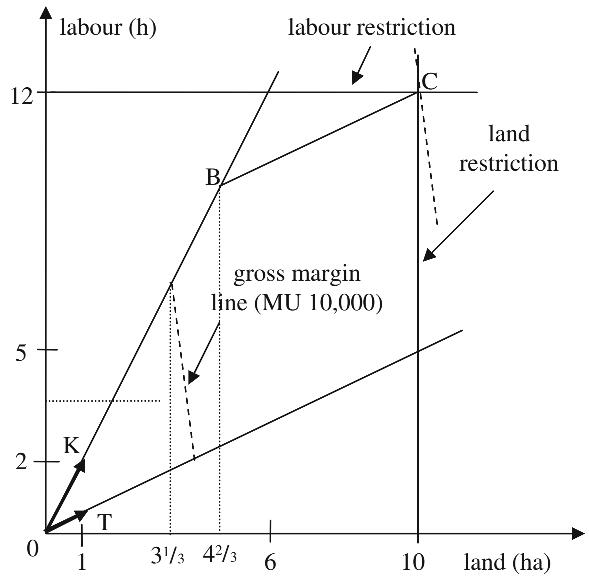
Hence, the analysis of the shadow price of land carried out here shows that the shadow price of land for the first 6 ha of land is MU 3,000 ha^{-1}. For the following 18 ha (up until 24 ha), it is MU 2,333. After this, it drops to MU 0 ha^{-1}.

A similar analysis could be carried out for labour. The reader should carry out her/his own calculations to verify that the following shadow prices of labour are correct: From 0 to 5 h: MU 5,000 h^{-1}. From 5 to 20 h: MU 333 h^{-1}. More than 20 h: MU 0 h^{-1}.

The graphical solution method is shown in Figs. 20.1 and 20.2 is the classical way of illustrating the solution to LP problems. This method provides a graphical illustration of the solution procedure—an illustration that can be useful as the basis for interpreting computer software-based solutions.

If there are more than two processes, the LP problem cannot be solved graphically as shown here. However, a graphical solution is still possible if there are only two restrictions (but it is fine with more than two processes).

Fig. 20.3 Optimisation of an LP problem



The method used in this case is to plot the fixed resources along the two axes of the coordinate system, and outline the processes as vectors in the plan, as shown in Fig. 20.3.

In Fig. 20.3, land is plotted along the horizontal axis, and labour is plotted along the vertical axis. The limited amounts are illustrated by the vertical line through the 10 ha point and the horizontal line through the 12 h point. Hence, an optimal solution is to be found within these lines.

The production processes are drawn as vectors. One unit of the production process for potatoes is drawn as the bold vector K (1 ha, 2 h, MU 3,000), and one unit of the production process for tomatoes is drawn as the bold vector T (1 ha, 0.5 h, MU 2,500).

The cultivation of the two crops is illustrated by the lines elongating the two vectors. As can be seen, it will be possible to cultivate 6 ha of potatoes, if the fixed resources are used only for potatoes, as the process line intersects the restriction line for labour at 6 h. If only tomatoes are grown, 10 ha can be cultivated.

As can be seen from the graphical solution shown in Fig. 20.1 above, growing one of the two crops is not optimal. It is, however, optimal to grow the two crops in combination.

To illustrate this, a line is inserted showing the crop combinations that generate the same gross margin (a gross margin line). E.g., 4 ha of tomatoes or $3^1/_3$ ha of potatoes generate a gross margin of MU 10,000. However, the linear combinations of these two points, as illustrated by the dotted line in Fig. 20.3, also generate a gross margin of MU 10,000.

Moving the gross margin line outwards in the plan generates a higher gross margin. Hence, a gross margin line that, e.g. is placed twice as far out on the plan as the one outlined in Fig. 20.3 illustrates points which all generate a gross margin of MU 20,000.

How far out can the gross margin line be shifted if it should still include the points within the two restriction lines for land and labour, respectively? As can be seen, the gross margin line leaves the restriction area at the exact point where it intersects point C. Hence, point C with a gross margin line going through this point illustrates the optimal cultivation plan.

Point C is a linear combination of the two process lines for potatoes and tomatoes, respectively. As can be seen, point C is precisely equal to a point that is produced by a combination of the potato vector 0B and the tomato vector BC. Hence, the optimal production consists of growing $4^2/_3$ ha of potatoes and $5^1/_3$ ha of tomatoes. (This solution is naturally equal to the solution found above in Fig. 20.1).

Solution with the Use of Computer Software

As shown here, LP problems can be solved graphically, if there are only two production processes or two restrictions. However, if there are more processes/ restrictions, the problem cannot be handled graphically and other methods must be used. The so-called simplex method is a solution algorithm that is generally applicable and the simplex method can, in principle, be used for calculation by hand. It is naturally easier to use computer software to solve LP problems. And there are many software programs on the market that can be used to solve LP problems, e.g. LINDO, GAMS, or SAS. The spreadsheet program Excel also contains an LP optimisation procedure. The aforementioned programs provide functionalities for generating the optimal solution as well as calculating the shadow prices and carrying out sensitivity analyses as described here.

The following contains a number of examples of formulations that will often be used for formulating planning problems within agriculture.

20.1.3 Process Formulation

The company's total production activities can be split up into a larger or smaller number of production processes. The first task when constructing an LP model for a company consists of splitting up the company into a planning-related number of processes corresponding to what can be called production branches, or operation branches. A suitable unit is chosen for each process, and the process' consumption of fixed resources per process unit is calculated. Finally, the parameters (the c's) that should be included in the criterion function are calculated. In most cases, this is simply the gross margin per process unit. However, for some processes this may consist of other values, which can be seen from the examples later.

According to the LP models' precondition about constant returns to scale, it is important that the processes are chosen in an expedient way. A process can be set up for what would normally be called a production branch, however, this can—if expedient—also be split up into several processes, just as two or more operation branches can be combined into one process. It could also be a matter of setting up more processes for the same production branch to include the possibility of modelling factor substitution in the model.

The unit for a process can either be a product or a factor unit. With regard to crop processes, it will often be natural to choose the process unit 1 ha. However, in greenhouse cultures for example, the unit 1 m^2 might be better. In connection with livestock processes, a livestock unit (1 cow, 1 sow, 1 finisher pig, 1 cow plus breeding, etc.) is often preferred; however, here it is also possible to choose factor units such as, e.g. 1 stable unit.

20.1.3.1 Production Processes

Following this general introduction, we will now look at some examples. As an introduction, an example of an LP matrix is shown in Table 20.2 which is followed by some comments below.

Processes for the sale of products

The process for growing seeds (A_1) is a "by the book" process with the gross margin as the objective function parameter and the technical coefficients being equal to the consumption of the limited resources, acreage and labour.

Processes for the intermediate products

Grass (A_4) and beets (A_5) are examples of intermediate products, i.e. products that are both produced and consumed in the model. The production processes for these two products have negative "gross margins" as "the gross margin" here only consists of the unit costs. The actual product value is not included, as the value of the two fodder items are generated by the process Cows (A_3), as each process unit (1 cow with breeding) consumes 2,500 FE (feed units) of grass and 1,700 FE of beets. The grass process (unit 1 ha) produces 7,000 FE ha $^{-1}$, and the process for beets (1 ha) produces 10,000 FE ha $^{-1}$. Please note that the right hand side for the two rows that control production of beets and grass is zero, corresponding to an initial storage of zero units. In reality, there might be a positive initial storage of, e.g. grass (in the form of grass silage stored from the last harvest). If this is the case, such an initial storage will be indicated on the right hand side. These rows ("Grass, storage" and "Beets, storage") control "intermediate products" and make sure that consumption does not exceed the produced amounts and the amount in storage.

Table 20.2 LP matrix with various production processes

	A_1 Seeds	A_2 Barley	A_3 Cows	A_4 Grass	A_5 Beets	A_6 Buy barley	A_7 Sell barley	Restriction	Right hand side
Gross margin	7,000	−2,000	14,000	−3,000	−6,000	−100	95	=	Max!
Acreage, hectares	1	1		1	1			≤	100
Cow stable, units			1					≤	120
Labour, h	15	10	55	16	35			≤	4,500
Grass, storage			2,500	−7,000				≤	0
Beets, storage			1,700		−10,000			≤	0
Barley, storage		−78	5			−1	1	≤	0

Processes for purchase and sale

Please note that the gross margin for barley (A_2) in Table 20.2 is also negative (MU −2,000). The reason is that, in this model formulation, barley is also considered as an "intermediate product" which can either be used as fodder for the cows, or sold at MU 95 hkg^{-1}. Hence, the gross margin for barley here also only includes the unit costs, as the revenue from the sale of the cereal crop is generated by the sales process (A_7) or (when used as fodder for the cows) through saved purchases of cereal (A_6) (at MU 100 hkg^{-1}) for feeding the cows. The barley process could also have been formulated in such a way that the sale of barley was included in the gross margin of barley [which would then have been MU 5,410 (check why!)]. And the process for the sale of barley (A_7) could then be deleted. If this is the case, the yield of −78 in the process for barley should also be removed. Please note that the initial storage of barley in the right hand side is equal to zero. Here, an initial storage greater than zero could also be inserted.

Processes for livestock

The process for cows can also easily be interpreted. The gross margin per cow (MU 14,000) is before the deduction of the variable costs for growing beets and grass, as these costs are deducted by the negative gross margin in the processes for beets and grass. Please note that the process A_3 uses the limited stable capacity of a total of 120 stable units, as well as the limited resource, labour. The process also uses cereal, beets, and grass which in this connection are also limited resources in that there is no storage (the right hand side is zero), and a possible consumption can therefore first be effected when these products are grown (cereal, beets, grass), or purchased (cereal).

Main and by-products

Table 20.3 contains an example of the production of barley where the main product is cereal which is sold and is valued at a gross margin of MU 5,000 ha^{-1}. However, apart from this, the production of barley contributes straw (5 tonnes ha^{-1}), which in this model can either be used as fodder for cows (300 kg per cow), or can be sold at a

Table 20.3 Example of processes with main and by-products as well as sale

	Barley, hectare A_1	Cows, piece A_2	Sale of straw, t A_3	Right hand side
Margin	5,000	12,000	250	=Max
Straw, storage	−5	0.3	1	≤0
Max sale			1	≤100

price of MU 250 tonne^{-1}. Furthermore, it contains the restriction that the sale of straw cannot exceed 100 tonnes (only the necessary rows and columns are included).

Multiple processes with the same product

It has previously been mentioned that the LP model has the unfortunate property of being based on a Leontief production function, which does not leave the possibility of substitution between inputs. Inputs are used in constant ratios, regardless of how many process units are included in the optimal solution.

This does not constitute a problem if the optimal ratio between the inputs is known in advance. However, as is well-known within general production economic theory, the optimal combination of inputs depends on the relative input prices. And with regard to fixed inputs, it can be difficult to know the actual price beforehand. These input prices (implicit prices) are equal to the shadow price of the restriction, and this shadow price is unknown until the model is solved.

An example of this kind of problem is illustrated in Table 20.4 which is similar to Table 20.2 but now with three processes for cows instead of one process. The process "Cows 1" is the same process as the cow process in Table 20.2, and this process presupposes that every cow is fed with 2,500 FE of grass and 1,700 FE of beets. However, there are other ways of feeding cows than by using this exact combination of grass and beets. Hence, it is also possible to use a smaller amount of beets and a larger amount of grass. And if grass turns out to be a cheaper fodder than beets, this solution would seem preferable. However, the price (the internal price—the shadow price) of grass (and beets) is not known until the model is solved. Therefore, it is not known beforehand whether a ratio corresponding to 2,500 FE of grass and 1,700 FE of beets is optimal.

This problem can be handled by formulating more, alternative processes for cows and then letting the actual model choose the optimal combination. In Table 20.4, a total of three different processes for cows have been formulated. The processes "Cows 2" and "Cows 3" use more grass than beets compared to "Cows 1". The consumption of labour per cow is also affected (less labour when using grass instead of beets as fodder), and the gross margin can also be affected (here, a somewhat smaller gross margin is presupposed when using more grass fodder).

With the aforementioned formulation, the model will itself choose the optimal process for cows, based on the internal prices (shadow prices) of grass and beets. The optimal solution might be a combination of two processes, as an LP model allows for linear combinations of processes.

20.1.3.2 Restrictions

As discussed in Chap. 19, restrictions in an LP model will primarily consist of the *fixed production factors*. The primary purpose of the LP model is to precisely allocate these fixed production factors in such a way that the highest possible profit

Table 20.4 Multiple processes for the same product

	A_1 Barley	A_2 Cows 1	A_3 Cows 2	A_4 Cows 3	A_5 Grass	A_6 Beets	Restriction	Right hand side
Gross margin	2,000	12,000	11,500	11,000	−3,000	−6,000	=	Max!
Acreage	1				1	1	≤	100
Cow stable		1	1	1			≤	120
Labour	10	55	50	45	16	35	≤	4,500
Grass, storage		2,500	3,000	3,300	−7,000		≤	0
Beets, storage		1,700	1,200	800		−10,000	≤	0

is achieved. It is important that all relevant restrictions are illustrated in the model for the planning model to provide a realistic illustration of a planning problem. The model should, on the other hand, not be constructed to be more restrictive than necessary, as the solution would then almost be given beforehand.

One of the key questions to be answered when constructing an LP planning model is, which factors are fixed and which are variable. There is no general answer to this question, as it will depend on the specific situation, including the length of the planning period especially.

The restriction system often includes factors that cannot be considered fixed but which are not actually variable either. As mentioned in Chap. 19, such factors are referred to as *conditional fixed* (or *conditional variable*) production factors. While fixed production factors are characterised economically by the sales price being zero, and the purchase price being infinite, the conditional variable production factors are characterised by the sales price being higher than zero, or the purchase price being smaller than infinite (as opposed to the variable factors where the purchase price is equal to the sales price). Situations may therefore arise in which it will be profitable to purchase or sell these normally fixed production factors, namely if the shadow price is very high (higher than the purchase price), or if the shadow price is very low (lower than the sales price). As these shadow prices are not known before the LP model is solved, it is useful to illustrate the possibility of a purchase or sale (or both) for these quasi-fixed factors in the model. In case of the shadow price being sufficiently high or low, the model itself would thus be able to decide whether to sell or buy.

Acreage restriction

The acreage of a given agricultural operation will, in most cases, be considered a fixed production factor. The purchase of land or, e.g. lease (renting) of land are often more long term decisions, which it is hardly appropriate to consider within the framework of a planning model that only includes one production period. Instead, the shadow price of the acreage restriction will often be of interest, as the shadow price indicates the maximum price that should be paid for the rent of land for this to be financially beneficial. The information about the shadow price of land provides a sound basis for making an offer for the purchase or rent of land.

If you wish to include the possibility of leasing land in an LP model this it can be done as shown in Table 20.5.

Table 20.5 Acreage restriction and lease of land

	Barley	Seeds	Beets	Lease of land	Restriction	Right hand side
Gross margin	6,000	7,000	−6,000	−5,000	=	Max!
Acreage	1	1	1	−1	≤	100
Max lease				1	≤	20

Table 20.6 Varying quality of land

	Barley	Wheat 1	Wheat 2	Grass only	Restriction	Right hand side
Gross margin	6,000	7,000	4,000	−1,500	=	Max!
Land type A	1	1			≤	80
Land type B			1		≤	20
Meadow				1	≤	10

As can be seen, there are already 100 ha that can be used for growing barley, seeds, or beets. However, apart from this, a "purchase process" provides the possibility of leasing land at MU 5,000 ha^{-1}. However, it is not possible to lease more than a maximum of 20 ha, which is outlined in the restriction rows "Max lease".

Similarly, the possibility of renting out land could also be illustrated. This would then be done by formulating a "sales process" for the rental of the land.

With regard to farm land, land could be of varying quality. Part of the land might only be suitable for certain crops (meadow and marsh for grass only) or there might be land types with different yield levels, e.g. where part of the land is sandy soil and part of the land is clayey soil.

Table 20.6 illustrates this type of variation.

In this example, the total acreage is 110 ha. However, part of it (20 ha) is of poor quality so that growing wheat only generates a gross margin of MU 4,000 versus MU 7,000 ha^{-1} on the good land (type A). The crop "grass only" with unit costs of MU 1,500 ha^{-1} is grown on the meadow which constitutes 10 ha.

Crop rotation restrictions

Certain crops require other crops as pre-crops. This is, e.g. true for grass seeds where the sowing of grass in the previous year is presupposed. Other crops (e.g. sugar beets and potatoes) may only be grown in a limited amount due to quotas and other restrictions. And finally, most crops have a tendency to generate a higher yield if another crop was grown on the acreage the previous year.

The restrictions mentioned here can be described under the umbrella concept of crop rotation restrictions and are illustrated in Table 20.7.

The top restriction is the "usual" acreage restriction which here indicates that the total acreage is 100 ha.

The second row indicates that cereal (barley and wheat) cannot be grown on more than 60 ha. The third row indicates that the "winter crops" winter oilseed rape and winter wheat cannot be grown on more than 50 ha. Such a restriction can, e.g. be due to the farmer not wanting to sow more than half the crops in autumn. The restriction concerning sugar beets is used if there is a quota of max 10 ha of sugar beets for the farm. The last two restrictions are related to the growing of grass seed. The first of these two restrictions indicates that first year grass seeds can only be grown if there, as a minimum, is a corresponding acreage of barley.

Table 20.7 Crop rotation restrictions

	Barley w/sowing grass	Winter wheat	First year grass seeds	Second year grass seeds	Sugar beets	Winter rape	Restriction	Right hand side
Gross margin	5,000	6,000	9,500	7,800	11,000	7,000	=	Max!
Acreage	1	1	1	1	1	1	≤	100
Max cereal	1	1					≤	60
Max Winter crop		1				1	≤	50
Max sugar beet					1		≤	10
Max grass seeds 1	−1		1				≤	0
Max grass seeds 2			−1	1			≤	0

The last restriction indicates that second year grass seeds can only be grown to the same extend as first year grass seeds (if there are no first year grass seeds, it is of course not possible to grow second year grass seeds!).

Please note that the "dynamic" relationships between the sowing of grass seeds and the growing of first and second year grass can be handled within the framework of a static LP model as shown here. And this will often be sufficient. However, it is important to be aware that the static model solution only provides a *long run equilibrium solution*. To be able to study a *development over time*, such dynamic (multi-period) problems should be formulated within the framework of an actual multi-period model, as discussed later in Sect. 20.2.

A particular type of crop rotation problem can be associated with the extent of growing of a crop. It is often the case that the larger the area of land cultivated with only one crop, the lower the average yield. This is either because the increased cultivation of one crop may take place on land of a poorer quality, with regard to the crop in question, or because with large acreages it is not possible to achieve yield advantages (at sustained growing) through crop rotation.

The creation of a model for diminishing yields with increasing acreage with one crop can take place as shown in Table 20.8.

The diminishing grain yield per hectare with an increased cultivation of barley is indicated in Table 20.8 by a decreasing gross margin per hectare. Cultivating up to 40 ha (of a total of 100 ha) generates a high gross margin (MU 5,000 ha^{-1}). If the acreage is increased to more than 40 ha, the grain yield decreases. The next 20 ha only generate MU 4,000 ha^{-1}, while the subsequent 20 ha generate MU 3,000 and, finally, the last 20 ha only generate MU 2,000 ha^{-1}. Illustrated as a production function with the yield of barley as a function of the supply of fixed input land, the relationship is as shown in Fig. 20.4.

As can be seen, a yield function with diminishing marginal returns is approximated by pieces of straight lines (here four) through linear combination of the four barley processes.

But is it possible to be certain that the LP model "activates" the four processes in the right order? How is it e.g. possible to be certain that the growing of 60 ha of barley takes place by use of the two processes Barley 1 and Barley 2, and not by use of the three processes Barley 2, Barley 3, and Barley 4? The answer is that it is

Table 20.8 Diminishing yields with increasing acreage

	Barley 1	Barley 2	Barley 3	Barley 4	Winter oilseed rape	Restriction	Right hand side
Gross margin	5,000	4,000	3,000	2,000	7,000	=	Max!
Acreage	1	1	1	1	1	≤	100
Max barley 1	1					≤	40
Max barley 2		1				≤	20
Max barley 3			1			≤	20
Max barley 4				1		≤	20
Max Winter oilseed rape					1	≤	50

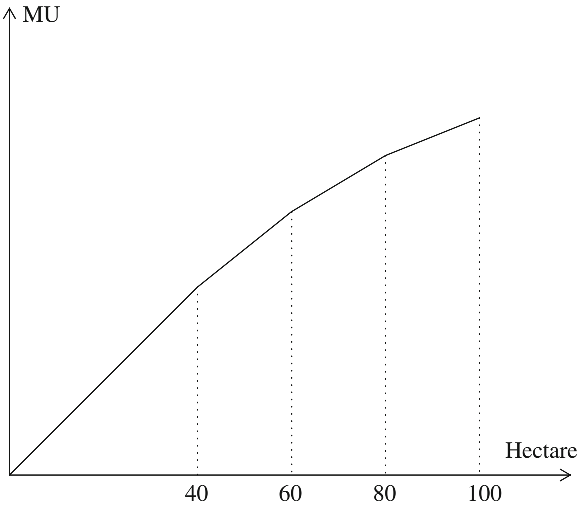


Fig. 20.4 Diminishing marginal returns with increasing acreage of barley

possible to be certain because the first combination of processes is the most profitable. Processes with a lower gross margin will only be included in the optimal solution when the processes with a higher gross margin have been fully used.

Even though the example did not show any problems regarding the process order when multiple processes should be used to illustrate non-linear shapes with piecewise linear functions, such problems might arise in other situations. We will return to such problems later during the discussion of integer programming in Sect. 20.3, as it is necessary to "control" the solution with the use of integer processes in such situations.

20.2 Multi-Period Models

20.2.1 Introduction

The ordinary LP model that has been used for production planning above is a static model. This means that the models parameters and the solution refer to a given point in time, or a given time period which in connection with production planning in agriculture will normally be a year.

If the planned activities are terminated fully within the relevant period (year), such a static (one-period model) is entirely sufficient for the calculation of the optimal production plan for the coming year. However, often the relevant activities have implications not only in the first year, but also in the subsequent year(s). If this is the case, it should be taken into consideration by including the long-term economic implications. After all, the aim of a company is normally to not only earn as much money in the first year, but to also earn money and survive in the coming years.

However, what is the problem then? Would it not be sufficient to only plan one year at a time, and then include the temporal aspect at the beginning of each new year by making a 1-year optimal plan with the static LP model?

Yes, under the condition that the individual years are independent in the way that the activities and the related profit are independent of the activities that were carried out in the previous year(s). However, particularly in agriculture, such independence is often not the case.

There are more examples of this. The fact that production takes time may imply that activities started in one period might not be terminated in the subsequent or even later periods (imagine, e.g. sowing grass or clover for seed production, the planting of blackcurrants, Christmas trees or strawberries, or the expansion of a cow stock by self-supply of calves). Also, the well-known agricultural pre-crop effect (including the fact that certain crops cannot be grown several years in a row on the same acreage) implies interdependence between the individual years. Investment activities in a given year also have an impact on the production possibilities several years ahead. Liquidity and capital restrictions may also have an impact. This is, e.g. the case when future activities are restricted by the inflow of capital acquired by means of savings in previous years, or taking out a loan where the payment of interest and instalments is distributed over a number of years.

In case of such dynamic relationships, the long-term implications should be considered when planning first year activities. As shown in the following, this is done by expanding the time frame, and thus the LP model, to include several years (time periods).

In the static LP model used above, it has to a certain extent been possible to consider these dynamic relationships. The sowing of grass seeds is, e.g. treated in the way that the model contained restrictions to ensure that the acreage with grass seeds could not exceed the size of the acreage with cover crop. Investments were also included and the economic implications were illustrated in the form of average yearly capital costs.

The problem with this kind of illustration is that the model solution is presumed to be applicable immediately from the beginning. On the other hand, there is no information about how to get from the present situation to the new production plan. If, e.g. the static LP model indicates that 20 ha of a given grass seed crop should be grown, and this should have been sowed last year, then the plan is of very little practical use, if this was in fact not done (should have been planned last year).

20.2.2 General Model

Based on the above, we will now look at how to adapt the general LP model to consider the dynamic relationships discussed here.

The idea is to construct a number of one-period LP models (one for each time period (year) in the future planning horizon). These one-period models are then

Table 20.9 Principle outline of a multi-period LP model

	Period 1	Period 2	....	Period T	Right hand side
Period 1	One-period LP model period 1				$\leq \mathbf{B}_1$
Period 2	$\mathbf{A}_{12}$	One-period LP model period 2			$\leq \mathbf{B}_2$
Period 3	$\mathbf{A}_{13}$	$\mathbf{A}_{23}$			$\leq \mathbf{B}_3$
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
Period *T*	$\mathbf{A}_{1T}$	$\mathbf{A}_{2T}$	....	One-period LP model Period *T*	$\leq \mathbf{B}_T$
All periods	$\mathbf{A}_{1A}$	$\mathbf{A}_{2A}$	....	$\mathbf{A}_{TA}$	$\leq \mathbf{B}_A$
Obj. func.	$\mathbf{C}_1$	$\mathbf{C}_2$	....	$\mathbf{C}_T$	Max!

"tied" together by means of processes and restrictions that move resources between each time period (year). This idea is illustrated in Table 20.9.

In the table, each of the "boxes" with a bold frame corresponds to an LP matrix for the period in question (hence, each of these "boxes" corresponds to a one-period LP model). The planning horizon consists of a total of *T* periods, and each period has a corresponding vector of restricting resources and the production-related restrictions $\mathbf{B}_1 \ldots \mathbf{B}_T$. Furthermore, the period as a whole may contain certain cross-period restrictions ($\mathbf{B}_A$). Finally, the last row contains an object function that can maximise some (linear) function of the total profit over all periods.

Apart from the original resources in each of the periods ($\mathbf{B}_1 \ldots \mathbf{B}_T$), there can be situations where the individual sub-periods "produce" resources that can be transferred to and used in the subsequent period(s). Hence, the matrix $\mathbf{A}_{ij}$ indicates the possibility of transferring resources—through the production activities in the period *i*—which will be available in the period *j*. The elements in these matrices will typically have a minus sign in front as this is a *delivery of* *resources* (or transfer of resources) for the subsequent period. In reality, this entails that to the original resources in, e.g. period 2 ($\mathbf{B}_2$) is added a resource quantity $\mathbf{A}_{12}$ which is transferred from period 1. The interpretation is the same for the other years.

With this type of model it is now possible to explicitly model the relationship between each period. The entire table in Table 20.9 is considered one big LP matrix with a right hand side that consists of a vector $\mathbf{B}$ which consists of the right hand sides of each of the years "piled up" on top of each other, and where the objective function maximises the total profit over all T time periods (e.g. net present value).

Mathematically, the multi-period LP model can be expressed as:

$$\text{Max Z} = \mathbf{C}_1\mathbf{X}_1 + \mathbf{C}_2\mathbf{X}_2 + \cdots + \mathbf{C}_t\mathbf{X}_t + \cdots + \mathbf{C}_T\mathbf{X}_T$$

subject to the constraints:

$$\begin{array}{ll}
\mathbf{A}_1\mathbf{X}_1 & \leq \mathbf{B}_1 \\
\mathbf{A}_{12}\mathbf{X}_1 + \mathbf{A}_2\mathbf{X}_2 & \leq \mathbf{B}_2 \\
\mathbf{A}_{13}\mathbf{X}_1 + \mathbf{A}_{23}\mathbf{X}_2 + \mathbf{A}_3\mathbf{X}_3 & \leq \mathbf{B}_3 \\
. & . \\
. & . \\
. & . \\
\mathbf{A}_{1T}\mathbf{X}_1 + \mathbf{A}_{2T}\mathbf{X}_2 + \mathbf{A}_{3T}\mathbf{X}_3 + \cdots + \mathbf{A}_{T-1,T}\mathbf{X}_{T-1} + \mathbf{A}_T\mathbf{X}_T & \leq \mathbf{B}_T
\end{array}$$

in which $\mathbf{C}_t$ is a vector of gross margins for period t, $\mathbf{X}_t$ is a vector of unknown process activities for period t, $\mathbf{A}_t$ is a matrix of technical coefficients for period t, $\mathbf{B}_t$ is a restriction vector for period t, and $\mathbf{A}_{tp}$ is a matrix that transfers resources from period t to period p.

If this is compared to the previously used one-period LP-model with the form:

$$\text{Max Z} = \mathbf{C}_1\mathbf{X}_1$$

subject to the constraints:

$$\mathbf{A}_1\mathbf{X}_1 \leq \mathbf{B}_1,$$

then it is evident that the multi-period LP model has many times more rows and columns than a one-period model. A multi-period LP model will therefore "swell up" easily, and a model with just ten periods is, in principle, ten times as big as a one-period model.

20.2.3 *Examples of Formulations*

In the following, several small examples of formulations of multi-period problems are illustrated. Most of the examples only include the processes and restrictions that are relevant for the illustration of how the problem model is structured.

20.2.3.1 Example 1: Crop Rotation

The first example in Table 20.10 includes a 3-year planning horizon where the farmer can grow barley (Barley), barley with the sowing of grass seed (Barley U), first year grass seed (Gr1), and second year grass seed (Gr2) in each of the years. The cultivation of first year grass seed presupposes the sowing of grass seed in barley (BarleyU) in the previous year, and growing second year grass seed presupposes, similarly, that first year grass seed (Gr1) has been cultivated in the previous year. All crops are indicated in units of 1 ha.

The available resources include 60 ha and 700 h of labour in each of the 3 years. Furthermore, at the beginning of year 1, grass seed has been sown on 5 ha the previous year (GrU1), and there are 7 ha with first year grass seed from the previous year (GrU2). For years 2 and 3, there is of course 0 ha of these two resource categories as grass seed has not yet been sown in year 1, just as no first year grass seed has been grown in year 1. The three vectors $\mathbf{B}_1$, $\mathbf{B}_2$, and $\mathbf{B}_3$ are, therefore, equal to [5, 7, 60, 700], [0, 0, 60, 700], and [0, 0, 60, 700], respectively.

If barley with the sowing of grass seed (BarleyU) is cultivated in year 1, this will provide a possibility of growing first year grass seed in year 2. The transfer of this resource from year 1 to year 2 is done by means of the coefficient −1 in the fifth row of the matrix. If first year grass (Gr1) is grown in year 1, this will provide a possibility of growing second year grass seed in year 2. The transfer of this resource from year 1 to year 2 is done by means of the coefficient −1 in the sixth row of the matrix. Thus, the transfer matrix $\mathbf{A}_{12}$ is equal to the sub-matrix with the two coefficients −1 and −1. The same form contains the transfer matrix $\mathbf{A}_{23}$. The three sub-matrices in the diagonal are $\mathbf{A}_1$, $\mathbf{A}_2$, and $\mathbf{A}_3$, respectively. Finally, the three vectors of gross margins are $\mathbf{C}_1$ = [4,800 3,900 5,700 5,400], $\mathbf{C}_2$ = [4,800 3,900 5,700 5,400], and $\mathbf{C}_3$ = [4,900, 4,100, 6,000 5,800].

In the example, the technical coefficients (labour consumption per hectare) are assumed to be the same during all 3 years. However, this is not necessarily a requirement. These coefficients may change from 1 year to the next and the model, therefore, contains the possibility of illustrating a potential technological development over time. The same is true for the gross margins, where any changes in prices and amounts over time can be illustrated at different gross margins for each of the years. In this example, a price increase and, thus, a higher gross margin is expected from year 2 to 3.

With regard to sub-periods of a length that will make it relevant to consider the alternative cost of capital, the payments should be added using the net present value formula. In the specific example, it is, e.g. possible to discount the gross margins for year 2 and year 3 to the net present value by multiplying the amount by $(1 + i)^{-1}$ and $(1 + i)^{-2}$, respectively, where i is the interest rate. Hence, the objective function expresses a net present value that is to be maximised.

Table 20.10 Sowing and growing perennial grass seed crops

	Year 1				Year 2				Year 3				Right hand side
	Bar-ley	Bar-leyU	Gr1	Gr2	Bar-ley	Bar-leyU	Gr1	Gr2	Bar-ley	Bar-leyU	Gr1	Gr2	
GrU1			1										≤5
GrU2				1									≤7
Acreage	1	1	1	1									≤60
Labour	10	11	9	11									≤700
GrU1		−1					1						≤0
GrU2			−1					1					≤0
Acreage					1	1	1	1					≤60
Labour					10	11	9	11					≤700
GrU1						−1					1		≤0
GrU2							−1					1	≤0
Acreage									1	1	1	1	≤60
Labour									10	11	9	11	≤700
Obj.	4,800	3,900	5,700	5,400	4,800	3,900	5,700	5,400	4,900	4,100	6,000	5,800	=max

Table 20.11 Expansion of a cow stock by self-supply

	Year 1				Year 2				Year 3				Right hand side
	Beets	Cows	Sale	Transfer	Beets	Cows	Sale	Transfer	Beets	Cows	Sale	Transfer	
Cows		1											≤50
Labour	50	70											≤1800
ControlT	−10	1											≤0
ControlC		−0.1	1	1									≤0
Cows				−1		1							≤50
Labour					50	70							≤1800
ControlT					−10	1							≤0
ControlC						−0.1	1	1					≤0
Cows								−1		1			≤50
Labour									50	70			≤1,800
ControlT									−10	1			≤0
ControlC										−0.1	1	1	≤0
Add.				1				1					≤10
Obj.	−6	12	7		−6	12	7		−6	12	7		=max

20.2.3.2 Example 2: Expansion of Stock of Cattle

Another example, concerning the expansion of a cow stock by the self-supply of heifers, is shown in Table 20.11.

As before, a planning horizon of 3 years is used. Initially, there is a stock of 50 cows. The farmer wants to expand the stock with a maximum of 10 cows over the 3 years (second last row). For this purpose, the "surplus production" of the farm's own heifers can be used. This surplus production consists of 0.1 heifers per full-year cow. These heifers can either by sold for MU 7,000 per unit ("Sale") or transferred to the cow stock as cows for the next year ("Transfer"). Due to restricted space in the table, only the rows and columns that are central to the illustration of this problem have been included. However, the growing of beets (yield = 10,000 FE ha $^{-1}$), and the consumption of the resource labour (max 1,800 h) have been included to symbolise an example of other processes and restrictions.

In this example, the transfer matrices $\mathbf{A}_{12}$ and $\mathbf{A}_{23}$ only contain one element, namely the number −1.

20.2.3.3 Example 3: Storage

In Table 20.12, a third example is shown. This concerns the production and sale of a product which should either be sold immediately, or stored and sold at a later date.

As can be seen, the production process "Prod." results in one unit of the product per process unit. The unit costs are MU 2 per unit. In the control row for period 1 (T1), it can be seen that there is an initial storage of 12 units, which together with the production from period 1, can either be sold (at MU 6 per unit) or stored for period 2. Storage entails a storage shrinkage of 5% and there is, therefore, only 0.95 units available for sale in period 2. Storage entails unit costs of MU 1 per unit. The sales price is MU 6 per unit in period 1, MU 7 per unit in period 2, and MU 6 per unit in period 3.

Apart from the illustrated rows and columns, the model might also contain other processes and restrictions. Please note that, as in the example in Table 20.11, the transfer matrices $\mathbf{A}_{12}$ and $\mathbf{A}_{23}$ only contain one element, namely the number −0.95.

Table 20.12 Storage of products

	Period 1			Period 2			Period 3			R HS
	Prod.	Sale	Storage	Sale	Storage	Prod.	Sale	Storage	Prod.	
T1	−1	1	1							≤12
T2			−0.95	−1	1	1				≤0
T3						−0.95	−1	1	1	≤0
Obj.	−2	6	−1	−2	7	−1	−2	6	−1	=max

20.2.3.4 Example 4: Investment and Financing

Table 20.13 shows an example of the financing of an investment by savings or through taking out a loan. Due to restrictions of space, only the rows and columns that are central to the illustration of this problem have been included.

The company's production is aggregated into one process "Produc.", which generates a payment contribution of MU c_j per process unit in year j. This production uses a_j units of a building capacity that is available initially in an amount of b units. The payment contributions can either be saved and transferred to next year at an interest rate of 9%, or be included in the financing of an investment.

It is now presupposed that it is possible to invest in additional building capacity of a total of 20 units. The investment amounts to MU 1,000. This investment can *either* be carried out in year 1, or year 2, or year 3. It takes one year to carry out the investment, and an investment initiated in year 1 does not provide the capacity (20 units) until year 2 (and 3).

The investment implies extra expenses for the operation of the plant of MU 15 year^{-1}. Furthermore, as indicated in the right hand side of the payment rows, there is a requirement for cash of MU 50 for the payment of the company's other fixed yearly expenses.

Investments can be financed through savings or through a loan. We presuppose that, in this case, a 10-year serial loan with an interest rate of 12% p.a. is taken out. Furthermore, a restriction of a maximum loan of MU 500 over the 3 years is presupposed.

The problem formulation is shown in Table 20.13.

In year 1, the loan process contributes MU 1 in cash per process unit [is included in the row of means of payment (Pay1)]. On the other hand, taking out a loan of MU 1 in year 1 means that instalment (MU 0.10) and interest (MU 0.12.) of a total of MU 0.22 should be paid in year 2, which has a negative effect on the payment row in year 2. In year 3, instalments of MU 0.10 and a 12% interest of MU 0.90 of the outstanding debt, i.e. a total of MU 0.21, should be paid. A similar interpretation can be found for the coefficients for the loan process in year 2.

The objective function calculates, in principle, the equity at the end of the 3-year period, as the residual value of the new investment (depreciation of MU 50 year^{-1}) minus the outstanding debt (MU 0.80 for loan taken out in year 1, MU 0.90 for loan taken out in year 2, and MU 1 for loan taken out in year 3). To this should be added interest to produce the three coefficients 0.90, 1.01, and 1.12 in the objective function.

20.2.4 Concluding Remarks

The multi-period LP model is particularly suited for illustrating a company's adjustment and development over time. The model type has previously been used by, e.g. Olsson (1970) and Renborg (1970) for illustrating problems of growth for

Table 20.13 Investment and financing

	Year 1				Year 2				Year 3				Right hand side
	Produc.	Saving	Invest	Loan	Produc.	Saving	Invest	Loan	Produc.	Saving	Invest	Loan	
Pay1	$-c_1$	1	1,000	−1									=−50
Build1	a_1												$\leq b$
Pay2		−1.09	15	0.22	$-c_2$	1	1,000	−1					=−50
Build2			−20		a_2								$\leq b$
Pay3			15	0.21		−1.09	15	0.22	$-c_3$	1	1,000	−1	=−50
Build3			−20				−20		a_3				$\leq b$
All years			1				1				1		≤1
Max loan				1				1				1	≤500
Obj.			900	−0.90			950	−1.01		1.09	1,000	−1.12	=max

farm companies. This model type has also been used by Köhne (1968) and Heidhues (1966) to illustrate investment and development planning problems for farm companies.

With regard to actual production planning, the model has primarily been used within greenhouse nurseries in Denmark, where the Dansk Erhvervsgartnerforening (DEG) has developed a multi-period model where each individual subperiod includes the individual weeks of the year. It seems also obvious to use a similar model type for optimising the within year production for outdoor nurseries.

The model would also have effective applications within other areas. Together with the Danish Meat Research Institute, Rasmussen (1992) developed a model for the optimisation of the utilisation of raw materials at pig slaughterhouses. A multi-period LP model with weeks (months) as sub-periods was used for this purpose.

20.3 Integer Programming

20.3.1 Introduction

The use of integer variables (integer processes) when solving LP problems, will often be a precondition for a realistic illustration of the planning problem in question. Introducing the requirement that some of the variables in the traditional LP model should be integers is called Mixed Integer Programming (MIP).

The following contains a number of examples of how integer variables can be used in the LP model.

The integer variables in MIP models can, in principle, be divided into two categories:

1. Natural integer variables.
2. Auxiliary integer variables.

Natural integer variables are variables that for natural reasons can only assume integer values in the solution. This could, e.g. be the decision about an investment, whereby the decision would be whether to invest in 0, 1, 2 or more units. For instance, it would be natural to declare a variable that measures the number of machines to buy to be an integer variable. However, even variables that are integer by nature need not always be declared as integer variables. In general, one should not use integer variables when it is unnecessary, because special software is needed and other problems are involved in their use (difficulties in the interpretation of shadow prices). A variable which assumes the value 115.4 in a continuous solution should not be declared an integer variable, even though the value should be an integer for the plan to be initiated, as the plan is very likely to be optimal even if the variable value is just rounded off.

The other type of integer variable is the one introduced artificially in the formulation of the planning problem to make the LP model more realistic. These variables make it possible to use LP for a number of problems that are not linear problems by nature, and problems which imply conditional restrictions. The integer variables are binary variables, i.e. integers that can only assume the values 0 or 1.

It is the use of the latter type of integer variable (0–1 variable or binary variable) that is discussed in the following.

The following symbols are used in the formulation examples:

y_i	indicates the integer (binary) variable (0–1 variable)
x_i	indicates a continuous variable
U_i	indicates an upper limit for variable no. i
cx_i, cy_i	indicates a criterion function value for variable no. i
b_j	indicates a right hand side element
a_{ij}	indicates an element in the coefficient matrix
RHS	indicates the restriction vector.

20.3.2 Examples of Formulations

20.3.2.1 Complementary Variables

If two continuous variables, x_1 and x_2, cannot be different from zero at the same time, it can be formulated by use of integer variables. Either one or the other (but not both) assume a level greater than zero.

Take as an example two fodder components, x_1 and x_2, that cannot be included in a fodder concentrate for a given animal at the same time. How is it possible to make sure that only one of these components is present at a time? (Table 20.14)

Comments:

1. The row “restriction” prevents y_1 and y_2 being both equal to 1. (Please note that this does not prevent them from both being equal to zero).
2. The row “consumption x_1” forces y_1 to be equal to 1, if $x_1 > 0$.
3. The row “consumption x_2” forces y_2 to be equal to 1, if $x_2 > 0$.

Table 20.14 Complementary variables

Rows	Variables				RHS
	x_1	x_2	y_1	y_2	
Restriction			1	1	≤ 1
Consumption x_1	-1		U_1		≥ 0
Consumption x_2		-1		U_2	≥ 0

20.3.2.2 Discrete Variables

Certain variables are only allowed to assume specific values, e.g. $y = 0$, 5, 15, or 30. A farmer is e.g. offered to enter into a contract for growing a specific crop, but it is only possible to enter into a contract of 0, 5, 15, or 30 ha. This can be formulated as shown in Table 20.15.

Comments:

1. The row "restriction" only allows one of the decision variables y_1, y_2, and y_3 to be equal to 1.
2. The row "choose" forces x_1 to assume a value corresponding to K_1, K_2, or K_3, depending on which integer variable, y_1, y_2, and y_3, is equal to 1.

20.3.2.3 Combination Variables

In the previous example, only one of several alternatives could be chosen. This example can be expanded to include the option of choosing several alternatives with a given set of alternatives.

A farmer, e.g. wants to only grow two out of three possible crops, but is indifferent as to the combination that can be chosen (Table 20.16).

Comments:

1. The row "restriction" implies that no more than two out of the three possible alternatives can be chosen. I.e. that only two of the variables y_1, y_2, and y_3 are equal to 1.
2. The rows "use x_1", "use x_2", and "use x_3" force the related decision variable to be equal to 1 if the continuous variable x_1, x_2, or x_3 is greater than 0.
3. Please note that this formulation can be expanded to include the choice of m alternatives among n possible alternatives.

Table 20.15 Discrete variables

Rows	Variables				RHS
	x_1	x_2	y_1	y_2	
Restriction		1	1	1	≤ 1
Choose	1	$-K_1$	$-K_2$	$-K_3$	=0

Table 20.16 Choice of combination of variables

Rows	Variables						RHS
	x_1	x_2	x_3	y_1	y_2	y_3	
Restriction				1	1	1	≤ 2
Use x_1	-1			U_1			≥ 0
Use x_2		-1			U_2		≥ 0
Use x_3			-1			U_3	≥ 0

20.3.2.4 Fixed Costs Problems (Set-Up Costs)

In a production process, the costs per produced unit are, e.g. constant (MU 10/unit, cx_1), but the production process requires a fixed cost of MU 500 (cy_1) for the production to be greater than 0 (Table 20.17).

Comments:

1. The row "invest" prevents x_1 from being different from zero before y_1 is equal to 1, so that a production cannot be carried out before an investment has been made.
2. U_1 is the capacity of the investment object, the size of the stable, the machine output per time unit, etc.

20.3.2.5 Investments that are Mutually Exclusive

A farmer can, e.g. choose between two stable types, y_1 and y_2. Stable type 1 costs MU 1,000,000, has a gross margin of MU 14,000 per cow, and requires a labour input of 40 h year^{-1}. Stable type 2 costs MU 1,500,000, has a gross margin of MU 16,000 per cow, and requires a labour input of 37 h year^{-1} (Table 20.18). Which stable should be built?

Comments:

1. The row "total" implies that either y_1 or y_2 is equal to 1, or both are equal to zero, but never that both are equal to 1.
2. The problem can be expanded to the choice of one or several options.

Connected Investments

It often happens that the implementation of a given investment implies that another investment should also be implemented.

Table 20.17 Start-up costs

Rows	Variables		RHS
	x_1	y_1	
Object function	$-cx_1$	$-cy_1$	=max
Invest	-1	U_1	≥ 0

Table 20.18 Investments that are mutually exclusive

Rows	Variables				RHS
	x_1	x_2	y_1	y_2	
Object function	cx_1	cx_2	$-cy_1$	$-cy_2$	=max
Choose y_1	-1		U_1		≥ 0
Choose y_2		-1		U_2	≥ 0
Total			1	1	≤ 1

Table 20.19 Associated investments

Rows	Variables				RHS
	x_1	x_2	y_1	y_2	
Object function	cx_1	cx_2	$-cy_1$	$-cy_2$	=max
Choose y_1	-1		U_1		≥ 0
Choose y_2		-1		U_2	≥ 0
Comb. y_2	-1			U_1	≥ 0

Table 20.20 Illustration of minimal production

Rows	Variables		RHS
	x_1	y_1	
Object function	cx_1	cy_1	=max
Min. x_1	1	-60	≥ 0
Calc. x_1	-1	U_1	≥ 0

If a farmer wants to invest in a beet harvester he/she will also have to invest in a tractor (if he/she does not already have a tractor with sufficient capacity for the harvester) (Table 20.19).

Comments:

1. The row "comb. y_2" makes sure that y_2 is equal to 1 (i.e. the investment is implemented) if x_1 is greater than zero.
2. Please note that this formulation implies that the investment y_2 can be implemented, without the need to implement investment y_1.

Minimal Production Scope

Often, the decision-maker will not be interested in small productions.

Examples:

1. A farmer wants to find out whether it is optimal to have cows and, if so, he will, e.g. have at least 60 cows.
2. Due to problems of weighing, a fodder concentrate company will not have components weighing less than, e.g. 50 kg (a sack) in its concentrate (Table 20.20).

Comments:

1. The row "min. x_1" implies that x_1 is at least 60 if $y_1 = 1$.
2. The row "calc. x_1" prevents x_1 from being different from zero, unless $y_1 = 1$.

20.3.2.6 Economies of Size

Increasing marginal productivity or an advantage of economies of size means that the financial result and/or resource consumption per process unit change with the process level (increasing financial result and decreasing resource consumption).

Fig. 20.5 Diminishing average costs (economies of size)

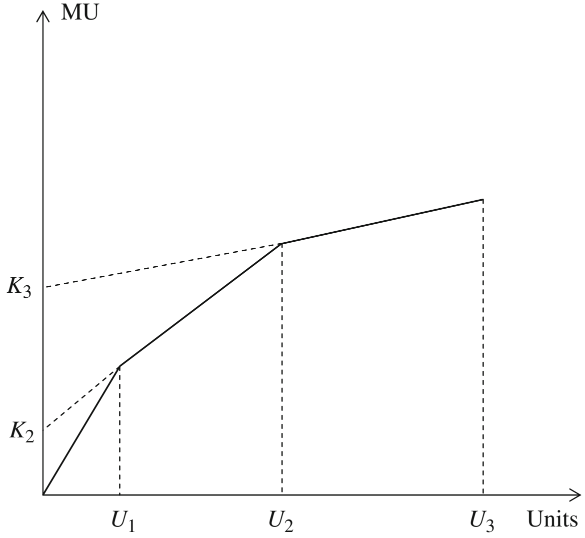


Table 20.21 Diminishing average costs

Rows	Variables						RHS
	x_1	X_2	x_3	y_1	y_2	y_3	
Object func.	cx_1	cx_2	cx_3				max
Process	a_{11}	a_{12}	a_{13}		K_2	K_3	$\leq b_i$
Comb-1	1			$-U_1$			≤ 0
Comb-2		1			$-U_2$		≤ 0
Comb-3			1			$-U_3$	≤ 0
Comb-4				1	1	1	≤ 1

Figure 20.5 shows a cost development with diminishing consumption per process unit at an increasing process level. This can, e.g. be a decreasing labour consumption per cow at an increased number of cows.

Comments:

1. The symbols in the formulation are shown in Fig. 20.5.
2. The row "comb-4" implies that only one of the integer processes y_1, y_2, and y_3 is equal to 1. This also implies that the rows: "comb-2", "comb-2", and "comb-3" are used to ensure that only one of the continuous processes is different from zero.
3. At a process level less than or equal to U_1, the row "comb-1" implies a resource consumption corresponding to the process x_1.
4. At a process level between U_1 and U_2, the row "comb-2" implies a resource consumption corresponding to the process $x_2 + K_2$ (the consumption K_2 corresponds to the consumption of x_1 at level U_1) (Table 20.21).

As it is not possible to use curved lines in MIP and/or LP models, it is necessary to make approximations by means of piecewise straight lines, as shown here. The approximation becomes better the more straight lines used.

20.3.3 Solving the Integer Programming Problems

LP problems with integer variables can be difficult to solve. One would expect this to simply be a question of first solving the problem without requiring an integer solution, and then rounding off the variables in question to the nearest integer solution.

This approach could be used with regard to solutions with variable values of a certain size. If an LP model based on continuous variables resulted in a solution of 50.6 units, it would probably not give rise to any major problems to simply round this number down to 50.

However, as can be seen from the examples above, integer programming is often used in situations where the allowed solutions consist of relatively small numbers, e.g. either 0 or 1. And in such cases, a solution procedure based on rounding off the number cannot be used.

The most commonly used method for solving mixed integer programming is the so-called branch and bound method (Schrage 1991). This method is, e.g. used in the LINDO computer software. Other computer software programs include functions for solving LP problems with integer variables. This is, e.g. true of the GAMS computer software.

The solution time is under all circumstances increased considerably when one or more of the variables are required to be integer variables. Also, the interpretation of shadow prices is more complicated.

References

Heidhues, T. (1966). Entwicklungsmöglichkeiten landwirtschaftlicher Betriebe unter verschiedenen Preisannahmen. *Berichte über Landwirtschaft, 181*, 141.

Köhne, M. (1968). Die Verwendung der linearen Programmierung zur Betriebsentwicklungsplanung in der Landwirtschaft. *Agrarwirtschaft, 25*, 113.

Olsson, R. (1970). *En modell för studier av jordbruksföretagets tillväkstproblem* (125 p. plus annex). Meddelanden från Institutionen för Ekonomie och Statistik, Landbrukshögskolan, Sweden.

Panell, D. J. (1997). *Introduction to Practical Linear Programming*. New York: Wiley.

Rasmussen, S. (1992). *The use of a multi-period LP-model as the core of a Decision Support System for a hog slaughterhouse*. Reports from Institut for Økonomi, Skov og Landskab, Samfundsvidenskabelig Serie, no. 3, KVL.

Renborg, U. (1970). Growth of the agricultural firm: Problems and theories. *Review of Marketing and Agricultural Economics,38*, 51–101.

Schrage, L. (1991). *LINDO—an optimization modelling system* (4th ed.). San Francisco: The Scientific Press.

Appendix: Profit Concepts

A.1 Introduction

The basic assumption in production economic theory is that the producer's objective is to maximise the company's revenue or profit. In connection with empirical analysis it is therefore important to ensure that there are well-defined guidelines for the calculation of revenue/profit from each individual branch of operation or the entire company.

As such, there seems to be no particular problems related to the calculation of the profit from each individual branch of operation or from the entire company. The profit (P) equals the total revenue (TR) minus the total costs (TC), i.e.:

$$P = TR - TC \tag{A.1}$$

In practice, the case is not that simple. In accounts statistics and in other connections there are various different profit measures of which some are relevant, while others are difficult to understand and without any actual informational value. The decisive point in this connection is that the purpose of the profit measure is often not completely clear. There is a tendency to confuse costs calculated according to the opportunity cost principle, and costs calculated according to the accounting principle, and it is often not made completely clear what the calculated profit measure should be used for.

Based on production economic theory, this chapter will look at the relevant profit measures and compare these with the profit measures that are used in practice.

A.2 The Purpose of Calculation of Profit

The calculation of profit at the company level can in principle have two purposes:

1. To function as the basis for an evaluation of the size of the amount that can be taken out of the company in terms of profit.
2. To function as the basis for planning (of future activities).

A.2.1 Profit from an Accounting Perspective

The aim here is to calculate the company profit. The calculation is carried out in the company accounts where the profit, in principle, is presented as the total revenue (gross output) minus costs. The total revenue is the value of the production during the period in question (normally a year). The costs are the estimated consumption of the input factors used in the production of the period measured in money terms.

Traditionally, the profit is calculated as the amount of money *left for the company owner*. To the extent that the owner has contributed to the production of the period (e.g. in the form of labour, management, risk taking, capital, etc.), the costs (as defined above) will thus *not* comprise the value of these input factors, and the profit will therefore be considered as the *remainder* for the remuneration for the owner's effort. In a sole proprietorship as e.g. within farming, this can be the farmers (and any unpaid family members) work effort, operational management and equity contribution. In private limited companies, it is the capital stock contribution plus the previous years' withheld dividends (reserves). The costs described in the first section should therefore rightly be defined as *the money value of input factors,* ***except the input factors which are at the owner's disposal***.

With the clarification introduced here, the ideal profit measure for the company can now be defined as the amount that the company owner (as remuneration for the effort rendered) can withdraw from the company without impairing the position of the company.

The formulation of this profit measure is relatively precise. It is based on the historical progress and the question to which an answer is sought is: If the company was completely re-established to its position at the start of the production period, how large an amount would remain for potential withdrawal—as profit—for the company owner.

The costs in the financial accounts are calculated according to *the accounting principle* as it is decided—for each of the production factors used—what it would cost to buy new or improve old ones for the company to be in the same condition as before.

The company accounts are normally prepared according to the following overall model:

Overview A.1 Accounting model

$$
\begin{aligned}
&\text{Gross output} - \text{unit costs}\\
&\quad = \text{gross margin} - \text{cash capacity cost}\\
&\quad = \text{earnings from operation} - \text{depreciations}\\
&\quad = \text{profit before interest} - \text{interest}\\
&\quad = \text{profit}
\end{aligned}
$$

The model for the calculation of profit shown here is based on the *acts* or *actions* of the company. Hence, the gross output and unit costs are associated with the *operational activities* of the company, and unit costs are related to the use of the resources in the ongoing production (costs related to the produced number of units). The *cash* capacity costs and the *calculated* capacity costs (depreciation) are associated with the *investment actions* of the company and are related to the maintenance of and investment in the company capacity. Finally, the item *interest* is associated with the *financing actions* of the company.

Deducting the owner's withdrawal from the calculated profit results in a remainder, i.e. the so-called *consolidation* which equals the operational contribution to the change in the company equity as follows:

Overview A.2 Calculation of consolidation

Profit – owner's withdrawal = consolidation (increase in equity)

Even though the method for calculating a company's (financial) profit seems a relatively well-defined and objective task, there is often a rather high degree of subjectivity associated with the calculations. This is basically due to the fact that it is not always entirely clear what is meant by profit, as the amount that can be withdrawn *without impairing the position the company*. Apart from requiring an interpretation of the actual meaning of *the position of the company*, two issues are especially relevant here. One concerns the importance of inflation. The other concerns the valuation of assets and liabilities.

There is general agreement that the financial description of the position of a company is expressed as *the company's equity*. The profit can then be defined as the amount that can be withdrawn from the company *without impairing its equity*.

With this definition in place, the two issues mentioned above will be described briefly in the following.

A.2.1.1 The Importance of Inflation

During inflation the purchasing power of money changes. When calculating the profit, it should therefore be decided whether the description "*without impairing its equity*" should refer to equity calculated in *current or fixed money values*.

Example A.1 A company has equity of MU 1,000,000 at the beginning of the year and equity of MU 1,300,000 at the end of the year. During the year there has been no withdrawal for the owner and the profit of the year is therefore MU 1,300,000–1,000,000 = MU 300,000.

However, during the same period there has been inflation of 6%. So as not to impair *the value in real terms* of the original equity of MU 1,000,000, an *adjustment of purchasing power* must be carried out, as the original equity must be adjusted to account for the inflation of 6% to maintain its value in real terms (purchasing power). The adjusted equity is then MU 1,060,000 and the profit of the year, based on the interpretation *the real value of the equity must not be impaired*, is therefore MU 240,000.

Hence, profit based on the interpretation that the *nominal value* of the equity must not be impaired will, in this example, result in a profit of MU 300,000. A profit based on the interpretation that the *real value* of the equity must not be impaired will result in a profit of MU 240,000. Which profit measure is the “right” one? And what profit measures should be used in practice?

The last questions could be dealt with by looking at how depreciation and interest are handled in the financial accounts. A profit measure based on the interpretation that *the real value of the equity* must be maintained would, namely, presuppose that depreciation is based on *replacement values*, and that the calculation of interest is based on a *real rate of return*.

In farming accounts, the calculation of depreciation was previously based on replacement values (the Danish Department of Farm Accounting and Management, the Danish Agricultural Advisory Centre (1999)). However, it has now been decided to base the calculation of depreciation on acquisition prices (the Danish Department of Farm Accounting and Management, the Danish Agricultural Advisory Centre (2001)). The calculation of interest costs in farming accounts has always been based on nominal interest.

According to the *Danish Financial Statements Act,* acquisition values should be depreciated and nominal interest should be deducted.

In practice, financial accounts are thus based on the principle that the company profit is the amount that can be withdrawn without impairing the *nominal* value of the equity.

A.2.1.2 Depreciation and Price Changes

Depreciation represents a consumption of resources and should in principle represent the period’s consumption of the assets that the company owns and utilises for two or more periods. This will typically be buildings and machinery with a limited lifetime where depreciation is used in an effort to distribute the costs related to the use of the asset over time.

If the price of the asset in question is constant over time, depreciation based on the acquisition principle and the reacquisition principle will give the same result. If an asset which was originally purchased for MU 100,000 and which is expected to be sold for MU 0 after 10 years, is depreciated over 10 years using the straight-line depreciation method, the amount depreciated would be the same each year (MU 10,000), regardless of whether the depreciation is based on the original acquisition price (MU 100,000) or the reacquisition price (which—due to the constant prices—is also MU 100,000[1]).

If, on the other hand, the price of the asset in question (or rather: the asset *type* in question) increases over time, depreciation based on the reacquisition principle will result in increased depreciation (and thereby—everything else being equal—a

[1] Please note that the use of constant prices here refers to constant prices measured in current monetary units (MU). If there is inflation (reduction of the purchasing power of money) at the same time, then this would be a fall in real prices.

lower profit (see overview A.1)) than when depreciating according to the acquisition principle.

Please consider the following example where the difference between the two depreciation principles is illustrated.

A company has just bought a machine for MU 100,000. The machine is expected to last 10 years after which it is to be sold for zero MU. Hence, the depreciation period is 10 years and the method is straight-line depreciation. The company has no debt (equity = total assets). *Earnings from operation* (see overview A.1) for the year are MU 60,500 which is found as the cash balance at the end of the year. Finally, the price of the machinery in question is presumed to increase by 5% each year so that a machine that was worth MU 100,000 at the start of the year will be worth MU 105,000 at the end of the year.

Depreciation according to the acquisition principle

The original acquisition price for machinery is MU 100,000. The depreciation for the year is therefore one tenth of MU 100,000, i.e. MU 10,000. If this amount is deducted from the earnings from operation of MU 60,500 for the year the resulting *profit of the year will be MU 50,500* (when using the model in overview A.1).

This amount could also have been calculated as the equity at the beginning of the year minus the equity at the end of the year. The equity at the beginning of the year is MU 100,000. At the end of the year it is MU 100,000 minus 10,000 in depreciation, plus MU 60,500 as the cash balance, i.e. MU 50,500. The difference of MU 50,500 in equity at the beginning of the year compared to the end of the year corresponds to the profit calculated above.

Depreciation according to the reacquisition principle

At the end of the year (when the financial accounts are prepared), the price of the *reacquisition* of a machine is determined. Due to a price increase, the reacquisition price has increased to MU 105,000 and the depreciation (according to the reacquisition principle) will therefore be one tenth of MU 105,000, i.e. MU 10,500. If this amount is deducted from the earnings from operation of MU 60,500 for the year, the resulting profit for the year will be MU 50,000 (when using the model in overview A.1). When depreciating according to the reacquisition principle, the resulting profit is (MU 500) lower than when depreciating according to the acquisition principle.

Now, to be certain and to carry out a kind of control, let us calculate the profit for the year as equity at the beginning of the year minus equity at the end of the year. The equity at the beginning of the year is MU 100,000. At the end of the year it is MU (1 + 0.05) = 105,000 minus MU 10,500 in depreciation, plus MU 60,500 as the cash balance, i.e. MU 155,000. Hence, the difference between equity at the beginning of the year compared to the end of the year is MU 55,000. However, it is *not* equal to the calculated profit of MU 50,000!!! What is the reason for this?

The reason is associated with the more general issue concerning the valuation of assets and liabilities, which will now be discussed.

A.2.1.3 The Implication of the Valuation of Assets and Liabilities

The equity can, as mentioned above, change without it being due to actual *consolidation, defined as profit minus withdrawal*. Changes in the equity can also be due to changes in the value of assets and liabilities—value changes that would normally not be included in the financial accounts. This can be illustrated by a simple example.

Example A.2 A company has calculated a profit of MU 300,000 for the year in its financial accounts (profit and loss account—see the accounting model in overview A.1). The owner has made no withdrawal. The equity could therefore be expected to show an increase of MU 300,000. However, the balance sheet shows equity that is MU 400,000 higher at the end of the year than at the beginning of the year. This increase is partly due to the profit of MU 300,000 and partly due to *an increase in the value of land by 100,000*. The value of the land, which at the beginning of the year was MU 2,000,000, has increased by 5% during the year and is thus valued at MU 2,100,000 in the balance sheet at the end of the year.

This example shows that it is not necessarily possible to equate changes in equity (adjusted with the owner's withdrawal, if any) with the financial profit. The reason for this is that value changes which are solely due to asset and liability price changes will normally (according to accounting conventions) not be included in the financial profit.

This means that a definition of the profit of a company as the amount that the owner withdraws without impairing the equity of the company does not necessarily correspond to the definition of a profit according to the normal accounting convention. In farming financial accounts, changes in the value of assets and liabilities are thus not included in the profit and loss account but are included as a separate "capital adjusting" item ("windfall changes") that is added directly to the equity.

It is not only within farming that value changes are not included in the company profit, as it is also the case in other industries. A profit that does not include the mentioned value changes is often referred to as an *operating profit*. Changes in the value of assets and liabilities are then included under *secondary items*, and the sum of primary and secondary items, less the owner's withdrawal if any, then constitutes the total change in equity.

A.2.1.4 Financial Indicators

According to the introduction to this chapter, the profit should cover the *remuneration for the input factors which are at the owner's disposal*.

With regard to public limited companies, the owner is the company's shareholders who have invested capital in the company. Therefore, the profit could be interpreted as the amount allocated for the return on the capital invested in the company by the shareholders.

With regard to sole proprietorships where both capital and labour is provided, a calculated remuneration for one of the two input factors can be deducted with the aim of calculating the remainder for the remuneration for the other.

If there is a particular focus on the return on capital, a *calculated remuneration* for *the owner's work effort* is deducted from the financial profit in overview A.1 as follows:

$$\text{Profit} - \text{calculated salary for the} = \text{Net income}$$

Then, the "Net income" is divided by equity, and the indicator "Return on equity" is calculated as follows:

$$\textbf{Return on equity} = \text{Net income/equity}$$

If, on the other hand, there is a particular focus on how the owner's work effort is being remunerated, a *calculated return* on the equity capital invested is deducted from the financial profit in overview A.1 as follows:

$$\text{Profit} - \text{calculated return of equity capital} = \text{Labour income for owner}$$

Then, the "Labour income for owner" is divided by the number of hours that the owner has worked in the company, and the result measure Labour income per hour for owner is calculated as follows:

$$\textbf{Labour income per hour for owner} = \\ \text{Labour income for owner/number of owner hours}$$

If you are interested in calculating *the return on total capital*, this is calculated as the so-called *rate of return* in the following way:

$$\textbf{Rate of return} = (\text{Profit before interest} - \text{calculated salary for the owner})/\text{Assets}$$

Hence, the rate of return is a measure of the return of the total invested capital—independent of the actual financing.

In a similar way, the remuneration for the *total* labour can be calculated as follows:

$$\begin{aligned} &\text{Profit} - \text{calculated return of capital} \\ &= \text{Labour income for} + \text{payroll cost according to financial accounts} \\ &= \text{Labour income} \end{aligned}$$

Then, the labour income per hour is calculated as follows:

$$\textbf{Labour income per hour} = \\ \text{Labour income/(number of owner hours + number of employee hours)}$$

Hence, the labour income per hour is an expression of how the total work effort is remunerated, independently of whether the work hour input comes from the owner or the employees.

The following examples illustrate these calculations.

Example A.3 A public limited company has assets amounting to MU 10,000,000 in total. The liabilities (loan capital) amount to MU 6,000,000 and the equity capital (share capital) then amounts to MU 4,000,000. MU 500,000 has been recognised as paid interest. The profit of the year is MU 800,000.

$$\begin{aligned}\text{Return on equity} &: 800{,}000/4{,}000{,}000 = 20\% \\ \text{Return on total capital} &: (800{,}000 + 500{,}000)/10{,}000{,}000 = 13\%\end{aligned}$$

Example A.4 A sole proprietorship has a total value of MU 8,000,000. The liabilities amount to MU 6,000,000 and the equity capital then amounts to MU 2,000,000. MU 500,000 has been recognised as paid interest. The owner has worked 1,200 hours in the company during the year, and MU 200 per hours or a total of MU 240,000 is included as remuneration for this. The profit of the year is MU 550,000.

$$\begin{aligned}\text{Return on equity} &: (550{,}000 - 240{,}000)/2{,}000{,}000 = 15.5\% \\ \text{Return on total capital} &: (550{,}000 + 500{,}000 - 240{,}000)/8{,}000{,}000 = 10.1\%\end{aligned}$$

The rate of return of equity indicates what the owner has achieved in return on the capital invested in the company. Whether a return of 15.5% can be considered high or low depends entirely on the return that could have been achieved by instead investing the money in assets with a corresponding uncertainty.

The return on total capital indicates what the return on the total capital has yielded in interest and is, thus, also a measure of interest to investors/potential investors within the industry.

Below is an example of the calculation of labour income.

Example A.5 A sole proprietorship has a total value of MU 8,000,000. The liabilities amount to MU 6,000,000 and the equity capital then amounts to MU 2,000,000. The owner has worked 1,200 hours in the company during the year. MU 200,000 has already been deducted under the cash capacity cost item as payment for 2,000 hours for a hired agricultural worker. The calculated return on equity capital is set at 4%. The profit for the year is MU 550,000.

$$\begin{aligned}&\text{Labour income for owner:} \\ &(550.000 - 0.04 \times 2{,}000{,}000)/1{,}200 = \text{MU } 350 \text{ per hour}\end{aligned}$$

$$\begin{aligned}&\text{Labour income for all labour:} \\ &(550.000 - 0.04 \times 2{,}000{,}000 + 200{,}000)/(1{,}200 + 2{,}000) = \text{MU } 209 \text{ per hour}\end{aligned}$$

The 4% calculated return on equity capital used in this example corresponds to the return that the Danish Institute of Food and Resource Economics has used in its statistics of farming accounts for a great many years. If the company's capital resources (assets) can be expected to increase with inflation, the percentage used should represent the real interest rate. Please refer to the subsequent courses in Investment Theory for an elaboration of this argument.

A.2.2 Calculation of Profit as the Basis for Planning

The aim here is to make decisions about the future operation of the company. This could include questions about choosing between alternative production branches or production methods, or questions regarding the evaluation of the total productivity of the company with regard to return on capital or labour remuneration, as the underlying question will be whether to invest the capital placed in the company in other activities instead, or to find other ways to apply the labour.

In connection with planning, costs are calculated according to *the opportunity cost principle*. This means that, for each of the planned input factors, an estimate would be made regarding any (better) alternative usage that might be relevant, and the earnings lost by *not* using the input factors in the alternative way is an expression of the costs.

The variable input factors can be bought and sold freely. Therefore, the (opportunity) cost is (normally) equal to the market price when purchasing/selling. The fixed input factors can neither be bought nor sold and the (opportunity) cost when using existing resources is therefore zero. This means that the net profit (P), which is equal to the total revenue (TR) minus variable costs (VC) minus fixed costs (FC) is equal to:

$$P = TR - VC - FC = TR - VC - 0 \equiv \text{GM} \tag{A.2}$$

in which GM stands for *gross margin*.

It is now completely clear why the gross margin (defined as total revenue minus variable costs) is such an important result measure in connection with planning. This is namely the relevant profit measure for planning and the opportunity cost principle is therefore used for the calculation of costs.

Please note that while the cost calculation in connection with the financial accounts ("history writing") is associated with how to *dispose of resources* (*unit costs* concern resources that have been planned to execute production activities, *capacity costs* concern resources that have been planned to create and maintain company capacity), the cost calculation in connection with planning is associated with the degree to which the resources are fixed. *Variable costs* concern the use of resources where the quantity can be varied if so desired, while *fixed costs* concern the use of resources where the quantity cannot be varied or there is no desire to vary the quantity within the relevant planning period.

When *planning operational activities*, the company's production capacity (plant) is considered as a fixed factor, and the related *capacity costs* will, in this connection, be considered as *fixed costs*. Similarly, *the unit costs* will, in this connection, be considered to be *variable costs*.

However, if the planning task is directed at *capacity actions*, then the company's capacity (or at least the part of it that the planning is directed at) will also be a variable factor, and the related *capacity costs* will therefore be *variable costs.*

The *gross margin*, defined as total revenue minus unit costs, is a very popular and important result measure in connection with planning, due to the fact that the important planning tasks in a company consist of *planning production activities*. In this connection, there is precise agreement between unit costs and variable costs, and the gross margin of the financial accounts becomes an important source of information as the basis for planning.

Please refer to the relevant postgraduate courses in financial management and information for a more thorough discussion of this comprehensive topic in company financial planning.

A.3 Conclusion

Ideally, a company's financial profit is calculated as the maximum amount that can be withdrawn without impairing the equity of the company. This profit measure is, however, not entirely unambiguously determined when there is inflation. In addition to this, the profit measures used in practice often deviate from the "ideal" measure, as price-related changes in the value of assets and liabilities are often not included in the financial profit.

The company's profit should cover the owner's effort. In sole proprietorships, this concerns labour and equity. In shareholder companies, the profit should cover the share capital return. To estimate the profitability, the rate of return and the labour income can be calculated.

In connection with planning, other cost and profit measures than the ones in the financial accounts are used. When planning the working/production activities of the company, the gross margin is a key result measure, and the financial accounts can be an important source of information for budgeting.

Index