Chapter 2

Review of technology and firms

The aim of this chapter is threefold. First, we shall introduce this book's vocabulary concerning firms' technology and technological change. Second, we shall refresh our memory of key notions from microeconomics relating to firms' behavior and factor market equilibrium under perfect competition. Finally, we outline a simple framework for the analysis of firms' behavior and factor market equilibrium under monopolistic competition.

The vocabulary pertaining to other aspects of the economy, for instance households' preferences and behavior, is better dealt with in close connection with the specific models to be discussed in the subsequent chapters. Regarding the distinction between discrete and continuous time analysis, the definitions contained in this chapter are applicable to both.

2.1 The production technology

Consider a two-factor production function given by

$$Y = F(K, L), \tag{2.1}$$

where Y is output (value added) per time unit, K is capital input per time unit, and L is labor input per time unit ($K \geq 0$, $L \geq 0$). We may think of (2.1) as describing the output of a firm, a sector, or the economy as a whole. It is in any case a very simplified description, ignoring the heterogeneity of output, capital, and labor. Yet, for many macroeconomic questions it may be a useful first approach. Note that in (2.1) not only Y but also K and L represent *flows,* that is, quantities per unit of time. If the time unit is one year, we think of K as measured in machine hours per year. Similarly, we think of L as measured in labor hours per year. Unless otherwise specified, it is understood that the rate of utilization of the production factors is constant over time and normalized to one for each production factor (cf. Section 1.2.2 of Chapter 1).

2.1.1 A neoclassical production function

By definition, K and L are non-negative. It is generally understood that a production function, $Y = F(K, L)$, is *continuous* and that $F(0, 0) = 0$ (no input, no output). Sometimes, when specific functional forms are used to represent a production function, that function may not be defined at points where $K = 0$ or $L = 0$ or both. In such a case we adopt the convention that the domain of the function is understood extended to include such boundary points whenever it is possible to assign function values to them such that continuity is maintained. For instance the function $F(K, L) = \alpha L + \beta KL/(K + L)$, where $\alpha > 0$ and $\beta > 0$, is not defined at $(K, L) = (0, 0)$. But by assigning the function value 0 to the point $(0, 0)$, we maintain continuity (and the "no input, no output" property).

We call the production function *neoclassical* if for all (K, L), with $K > 0$ and $L > 0$, the following additional conditions are satisfied:

(a) $F(K, L)$ has continuous first- and second-order partial derivatives satisfying:

$$F_K > 0, \quad F_L > 0, \tag{2.2}$$

$$F_{KK} < 0, \quad F_{LL} < 0. \tag{2.3}$$

(b) $F(K, L)$ is strictly quasiconcave (i.e., the level curves, also called isoquants, are strictly convex to the origin).

In words: (a) says that a neoclassical production function has continuous substitution possibilities between K and L and the *marginal productivities* are positive, but diminishing in own factor. Thus, for a given number of machines, adding one more unit of labor, adds to output, but less so, the higher is already the labor input. And (b) says that every isoquant, $F(K, L) = \bar{Y}$, has a form qualitatively similar to that shown in Fig. 2.1.[1] When we speak of for example F_L as the marginal *productivity* of labor, it is because the "pure" partial derivative, $\partial Y/\partial L = F_L$, has the denomination of a productivity (output units/yr)/(man-yrs/yr). It is quite common, however, to refer to F_L as the marginal *product* of labor. Then a unit marginal increase in the labor input is understood: $\Delta Y \approx (\partial Y/\partial L)\Delta L = \partial Y/\partial L$ when $\Delta L = 1$.

[1] A refresher on mathematical terms such as *boundary point, convex function,* etc. is contained in Math Tools at the end of this book.

Similarly, F_K can be interpreted as the marginal *productivity* of capital or as the marginal *product* of capital. In the latter case it is understood that $\Delta K = 1$, so that $\Delta Y \approx (\partial Y/\partial K)\Delta K = \partial Y/\partial K$.

The definition of a neoclassical production function can be extended to the case of n inputs, in the amounts $X_1, X_2, \ldots, X_n$, that is, $Y = F(X_1, X_2, \ldots, X_n)$. Then F is called neoclassical, if F is strictly quasiconcave and all the marginal productivities are positive, but diminishing.

Returning to the two-factor case, since $F(K, L)$ presumably depends on the level of technical knowledge and this level depends on time, t, we might want to replace (2.1) by

$$Y_t = F^t(K_t, L_t), \tag{2.4}$$

where the superscript on F indicates that the production function may shift over time, due to changes in technology. We then say that $F^t(\cdot)$ is a neoclassical production function if it satisfies the conditions (a) and (b) for all pairs (K_t, L_t). *Technical progress* can then be said to occur when, for K_t and L_t held constant, output increases with t. For convenience, to begin with we skip the explicit reference to time and level of technology.

The marginal rate of substitution Given a neoclassical production function F, we consider the isoquant defined by $F(K, L) = \bar{Y}$, where $\bar{Y}$ is a positive constant. The *marginal rate of substitution*, MRS_{KL}, of K for L at the point (K, L) is defined as the absolute slope of the isoquant at that point, cf. Fig. 2.1. The equation $F(K, L) = \bar{Y}$ defines K as an implicit function of L. By implicit differentiation we find $F_K(K, L)dK/dL + F_L(K, L) = 0$, from which follows

$$MRS_{KL} = -\frac{dK}{dL}_{|Y=\bar{Y}} = \frac{F_L(K, L)}{F_K(K, L)} > 0. \tag{2.5}$$

That is, MRS_{KL} measures the amount of K that can be saved (approximately) by applying an extra unit of labor. In turn, this equals the ratio of the marginal productivities of labor and capital, respectively.[2] Since F is neoclassical, by definition F is strictly quasi-concave and so the marginal rate of substitution is diminishing as substitution proceeds, i.e., as the labor input is further increased along a given isoquant. Notice that this feature characterizes the marginal rate of substitution for any neoclassical production function, whatever the returns to scale (see below).

When we want to draw attention to the dependency of the marginal rate of substitution on the factor combination considered, we write $MRS_{KL}(K, L)$.

[2]The subscript $|Y = \bar{Y}$ in (2.5) indicates that we are moving along a given isoquant, $F(K, L) = \bar{Y}$.

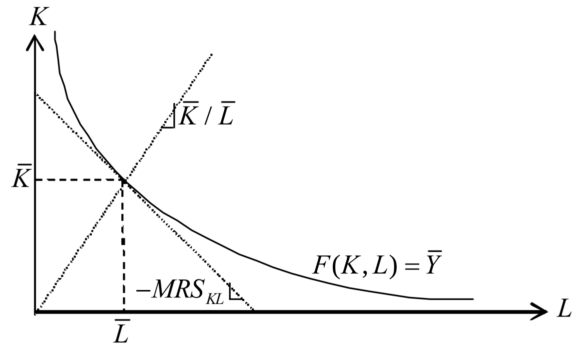

Figure 2.1: MRS_{KL} as the absolute slope of the isoquant.

Sometimes in the literature, the marginal rate of substitution between two production factors, K and L, is called the *technical* rate of substitution in order to distinguish from a consumer's marginal rate of substitution between two consumption goods.

As is well-known from microeconomics, a firm that minimizes production costs for a given output level and given factor prices, will choose factor combination such that MRS_{KL} equals the ratio of the factor prices. If $F(K, L)$ is homogeneous of degree q, then the marginal rate of substitution depends only on the factor proportion and is thus the same at any point on the ray $K = (\bar{K}/\bar{L})L$. That is, in this case the expansion path is a straight line.

The Inada conditions A neoclassical production function is said to satisfy the *Inada conditions*[3] if

$$\lim_{K\to 0} F_K(K,L) = \infty, \lim_{K\to\infty} F_K(K,L) = 0, \tag{2.6}$$
$$\lim_{L\to 0} F_L(K,L) = \infty, \lim_{L\to\infty} F_L(K,L) = 0. \tag{2.7}$$

In this case, the marginal productivity of either production factor has no upper bound when the input of the factor becomes infinitely small. And the marginal productivity is vanishing when the input of the factor increases without bound. Actually, (2.6) and (2.7) express *four* conditions, which it is preferable to consider separately and label one by one. In (2.6) we have two *Inada conditions for* MPK (the marginal productivity of capital), the first being a *lower*, the second an *upper* Inada condition for MPK. And in (2.7) we have two *Inada conditions for* MPL (the marginal productivity of labor), the first being a *lower*, the second an *upper* Inada condition for MPL. In the

[3] After the Japanese economist Ken-Ichi Inada, 1925-2002.

literature, when a sentence like "the Inada conditions are assumed" appears, it is sometimes not made clear which, and how many, of the four are meant. Unless it is evident from the context, it is better to be explicit about what is meant.

The definition of a neoclassical production function we gave above is quite common in macroeconomic journal articles and convenient because of its flexibility. There are economic growth textbooks that define a neoclassical production function more narrowly by including the Inada conditions as a requirement for calling the production function neoclassical. In contrast, when in a given context we need one or another Inada condition, we state it explicitly as an additional assumption.

2.1.2 Returns to scale

If all the inputs are multiplied by some factor, is output then multiplied by the same factor? There may be different answers to this question, depending on circumstances. We consider a production function $F(K, L)$ where $K > 0$ and $L > 0$. Then F is said to have *constant returns to scale* (CRS for short) if it is homogeneous of degree one, i.e., if for all (K, L) and all $\lambda > 0$,

$$F(\lambda K, \lambda L) = \lambda F(K, L).$$

As all inputs are scaled up or down by some factor, output is scaled up or down by the same factor.[4] The assumption of CRS is often defended by the *replication argument.* Before discussing this argument, lets us define the two alternative "pure" cases.

The production function $F(K, L)$ is said to have *increasing returns to scale* (IRS for short) if, for all (K, L) and all $\lambda > 1$,

$$F(\lambda K, \lambda L) > \lambda F(K, L).$$

That is, IRS is present if, when all inputs are scaled up by some factor, output is scaled up by *more* than this factor. The existence of gains by specialization and division of labor, synergy effects, etc. sometimes speak in support of this assumption, at least up to a certain level of production. The assumption is also called the *economies of scale* assumption.

Another possibility is *decreasing returns to scale* (DRS). This is said to occur when for all (K, L) and all $\lambda > 1$,

$$F(\lambda K, \lambda L) < \lambda F(K, L).$$

[4] In their definition of a neoclassical production function some textbooks add constant returns to scale as a requirement besides (a) and (b) (and perhaps the Inada conditions). Our terminology is different; when in a given context we need an assumption of constant returns to scale, we state it as an additional assumption.

That is, DRS is present if, when all inputs are scaled up by some factor, output is scaled up by *less* than this factor. This assumption is also called the *diseconomies of scale* assumption. The underlying hypothesis may be that control and coordination problems confine the expansion of size. Or, considering the "replication argument" below, DRS may simply reflect that behind the scene there is an additional production factor, for example land or a irreplaceable quality of management, which is tacitly held fixed, when the factors of production are varied.

EXAMPLE 1 The production function

$$Y = AK^{\alpha}L^{\beta}, \qquad A > 0, 0 < \alpha < 1, 0 < \beta < 1, \tag{2.8}$$

where A, α, and β are given parameters, is called a *Cobb-Douglas production function*. The parameter A depends on the choice of measurement units; for a given such choice it reflects the "total factor productivity". Exercise 2.2 asks the reader to verify that (2.8) satisfies (a) and (b) above and is therefore a neoclassical production function. The function is homogeneous of degree $\alpha + \beta$. If $\alpha + \beta = 1$, there are CRS. If $\alpha + \beta < 1$, there are DRS, and if $\alpha + \beta > 1$, there are IRS. Note that α and β must be less than 1 in order not to violate the diminishing marginal productivity condition. □

EXAMPLE 2 The production function

$$Y = \min(AK, BL), \qquad A > 0, B > 0, \tag{2.9}$$

where A and B are given parameters, is called a *Leontief production function* or a *fixed-coefficients production function;* A and B are called the *technical coefficients*. The function is not neoclassical, since the conditions (a) and (b) are not satisfied. Indeed, with this production function the production factors are not substitutable at all. This case is also known as the case of *perfect complementarity*. The interpretation is that already installed production equipment requires a fixed number of workers to operate it. The inverse of the parameters A and B indicate the required capital input per unit of output and the required labor input per unit of output, respectively. Extended to many inputs, this type of production function is often used in multi-sector input-output models (also called Leontief models). In aggregate analysis neoclassical production functions, allowing substitution between capital and labor, are more popular than Leontief functions. But sometimes the latter are preferred, in particular in short-run analysis with focus on the use of already installed equipment where the substitution possibilities are limited. As (2.9) reads, the function has CRS. A generalized form of the Leontief function is $Y = \min(AK^{\gamma}, BL^{\gamma})$, where $\gamma > 0$. When $\gamma < 1$, there are DRS, and when $\gamma > 1$, there are IRS. □

The replication argument The assumption of CRS is widely used in macroeconomics. The model builder may appeal to the *replication argument* saying that by, conceptually, doubling all the inputs, we should always be able to double the output, since we just "replicate" what we are already doing. One should be aware that the CRS assumption is about *technology* $-$ limits to the availability of resources is another question. The CRS assumption and the replication argument presuppose that *all* the relevant inputs are explicit as arguments in the production function and that these are changed equiproportionately. Concerning our present production function $F(\cdot)$, one could easily argue that besides capital and labor, also land is a necessary input and should appear as a separate argument.[5] Then, on the basis of the replication argument we should in fact expect DRS wrt. capital and labor alone. In manufacturing and services, empirically, this and other possible sources for departure from CRS may be minor and so many macroeconomists feel comfortable enough with assuming CRS wrt. K and L alone, at least as a first approximation. This approximation is, however, less applicable to poor countries, where natural resources may be a quantitatively important production factor and an important part of national wealth.

Another problem with the replication argument is the following. The CRS claim is that by changing all the inputs equiproportionately by any positive factor λ, which does not have to be an integer, the firm should be able to get output changed by the same factor. Hence, the replication argument requires that indivisibilities are negligible, which is certainly not always the case. In fact, the replication argument is more an argument against DRS than *for* CRS in particular. The argument does not rule out IRS due to synergy effects as size is increased.

Sometimes the replication line of reasoning is given a more precise form. This gives occasion for introducing a useful local measure of returns to scale.

The elasticity of scale To allow for indivisibilities and mixed cases (for example IRS at low levels of production and CRS or DRS at higher levels), we need a local measure of returns to scale. One defines the *elasticity of scale*, $\eta(K, L)$, of F at the point (K, L), where $F(K, L) > 0$, as

$$\eta(K,L) = \frac{\theta}{F(K,L)}\frac{dF(\theta K,\theta L)}{d\theta} = \frac{dF(\theta K,\theta L)/F(K,L)}{d\theta/\theta}, \text{ evaluated at } \theta = 1. \tag{2.10}$$

[5] Recall from Chapter 1 that we think of "capital" as producible means of production, whereas "land" refers to non-producible natural resources, including for example building sites. If an industrial firm decides to duplicate what it has been doing, it needs a piece of land to build another plant like the first.

So the elasticity of scale at a point (K, L) indicates the (approximate) percentage increase in output when both inputs are increased by 1 per cent. We say that

$$\text{if } \eta(K,L) \begin{cases} >1, & \text{then there are locally } \textit{IRS}, \\ =1, & \text{then there are locally } \textit{CRS}, \\ <1, & \text{then there are locally } \textit{DRS}. \end{cases} \tag{2.11}$$

The production function *may* have the same elasticity of scale everywhere. This is the case if and only if the production function is homogeneous. If F is homogeneous of degree h, then $\eta(K, L) = h$ and h is called the *elasticity of scale parameter.*

Note that the elasticity of scale at a point (K, L) will always equal the sum of the partial output elasticities at that point:

$$\eta(K,L) = \frac{F_K(K,L)K}{F(K,L)} + \frac{F_L(K,L)L}{F(K,L)}. \tag{2.12}$$

This follows from the definition in (2.10) by taking into account that

$$\begin{aligned} \frac{dF(\theta K, \theta L)}{d\theta} &= F_K(\theta K, \theta L)K + F_L(\theta K, \theta L)L \\ &= F_K(K,L)K + F_L(K,L)L, \text{ when evaluated at } \theta = 1. \end{aligned}$$

Fig. 2.2 illustrates a popular case from microeconomics, a U-shaped average cost curve from the perspective of the individual firm (or plant): at low levels of output there are falling average costs (thus IRS), at higher levels rising average costs (thus DRS). Given the input prices, w_K and w_L, and a specified output level, $\bar{Y}$, we know that the cost minimizing factor combination $(\bar{K}, \bar{L})$ is such that $F_L(\bar{K}, \bar{L})/F_K(\bar{K}, \bar{L}) = w_L/w_K$. It is easy to show that the elasticity of scale at $(\bar{K}, \bar{L})$ will satisfy:

$$\eta(\bar{K}, \bar{L}) = \frac{LAC(\bar{Y})}{LMC(\bar{Y})}, \tag{2.13}$$

where $LAC(\bar{Y})$ is average costs (the minimum unit cost associated with producing $\bar{Y}$) and $LMC(\bar{Y})$ is marginal costs at the output level $\bar{Y}$ (see Appendix A). The L in LAC and LMC stands for "long-run", indicating that both capital and labor are considered variable production factors within the period considered. At the optimal plant size, Y^*, there is equality between LAC and LMC, implying a unit elasticity of scale, that is, locally we have CRS.

This provides a more subtle replication argument for CRS at the aggregate level. Even though technologies may differ across firms, the surviving firms

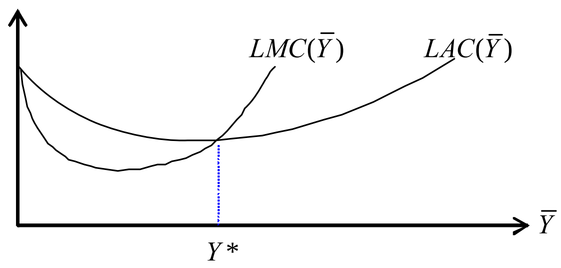

Figure 2.2: Locally CRS at optimal plant size.

in a competitive market will have the same average costs at the optimal plant size. In the medium and long run, changes in aggregate output will take place primarily by entry and exit of optimal-size plants. Then, with a large number of relatively small plants, each producing at approximately constant unit costs for small output variations, we can without substantial error assume constant returns to scale at the aggregate level. So the argument goes. Notice, however, that even in this form the replication argument is not entirely convincing since the question of indivisibility remains. The optimal plant size may be large relative to the market − and is in fact so in many industries. Besides, in this case also the perfect competition premise breaks down.

The empirical evidence concerning returns to scale is mixed (see the literature notes at the end of the chapter). Notwithstanding the theoretical and empirical ambiguities, the assumption of CRS wrt. capital and labor has a prominent role in macroeconomics. In many contexts it is regarded as an acceptable approximation and a convenient simple background for studying the question at hand.

2.1.3 Properties of the production function under CRS

Expedient inferences of the CRS assumption include:

(i) marginal costs are constant and equal to average costs (put $\eta(\bar{K}, \bar{L}) \equiv 1$ in (2.13));

(ii) if production factors are paid according to their marginal productivities, factor payments exactly exhaust total output so that pure profits are neither positive nor negative (put $\eta(K, L) \equiv 1$ in (2.12));

(iii) a production function known to exhibit CRS and satisfy property (a) from the definition of a neoclassical production function above, will automatically satisfy also property (b) and consequently *be* neoclassical (see Appendix A);

(iv) a neoclassical two-factor production function with CRS has always $F_{KL} > 0$, i.e., it exhibits "gross-complementarity" between K and L;

(v) a two-factor production function, known to have CRS and be twice continuously differentiable with positive marginal productivity of each factor everywhere in such a way that all isoquants are strictly convex to the origin, *must* have *diminishing* marginal productivities everywhere.[6]

A principal implication of the CRS assumption is that it allows a reduction of dimensionality. Considering a neoclassical production function, $Y = F(K, L)$ with $L > 0$, we can under CRS write $F(K, L) = LF(K/L, 1) \equiv Lf(k)$, where $k \equiv K/L$ is the *capital intensity* and $f(k)$ is the *production function in intensive form* (sometimes named the per capita production function). Thus output per unit of labor depends only on the capital intensity:

$$y \equiv \frac{Y}{L} = f(k).$$

When the original production function F is neoclassical, under CRS the expression for the marginal productivity of capital simplifies:

$$F_K(K, L) = \frac{\partial Y}{\partial K} = \frac{\partial \left[Lf(k) \right]}{\partial K} = Lf'(k)\frac{\partial k}{\partial K} = f'(k). \tag{2.14}$$

And the marginal productivity of labor can be written

$$\begin{aligned} F_L(K, L) &= \frac{\partial Y}{\partial L} = \frac{\partial \left[Lf(k) \right]}{\partial L} = f(k) + Lf'(k)\frac{\partial k}{\partial L} \\ &= f(k) + Lf'(k)K(-L^{-2}) = f(k) - f'(k)k. \end{aligned} \tag{2.15}$$

A neoclassical CRS production function in intensive form always has a positive first derivative and a negative second derivative, i.e., $f' > 0$ and $f'' < 0$. The property $f' > 0$ follows from (2.14) and (2.2). And the property $f'' < 0$ follows from (2.3) combined with

$$F_{KK}(K, L) = \frac{\partial f'(k)}{\partial K} = f''(k)\frac{\partial k}{\partial K} = f''(k)\frac{1}{L}.$$

For a neoclassical production function with CRS, we also have

$$f(k) - f'(k)k > 0 \text{ for all } k > 0, \tag{2.16}$$

[6] Proof of claim (iii) is in Appendix A and proofs of claim (iv) and (v) are in Appendix B.

as well as

$$\lim_{k\to 0}\left[f(k) - f'(k)k\right] = f(0). \qquad (2.17)$$

Indeed, from the mean value theorem[7] we know there exists a number $a \in (0,1)$ such that for any given $k > 0$ we have $f(k)-f(0) = f'(ak)k$. From this follows $f(k)-f'(ak)k = f(0) < f(k)-f'(k)k$, since $f'(ak) > f'(k)$ by $f'' < 0$. In view of $f(0) \geq 0$, this establishes (2.16). And from $f(k) > f(k) - f'(k)k > f(0)$ and continuity of f follows (2.17).

Under CRS the Inada conditions for MPK can be written

$$\lim_{k\to 0} f'(k) = \infty, \qquad \lim_{k\to \infty} f'(k) = 0. \qquad (2.18)$$

An input which must be positive for positive output to arise is called an *essential input*. The second part of (2.18), representing the upper Inada condition for MPK under CRS, has the implication that *labor* is an essential input; but capital need not be, as the production function $f(k) = a+bk/(1+k)$, $a > 0, b > 0$, illustrates. Similarly, under CRS the upper Inada condition for MPL implies that *capital* is an essential input. These claims are proved in Appendix C. Combining these results, when *both* the upper Inada conditions hold and CRS obtains, then both capital and labor are essential inputs.[8]

Fig. 2.3 is drawn to provide an intuitive understanding of a neoclassical CRS production function and at the same time illustrate that the lower Inada conditions are more questionable than the upper Inada conditions. The left panel of Fig. 2.3 shows output per unit of labor for a CRS neoclassical production function satisfying the Inada conditions for MPK. The $f(k)$ in the diagram could for instance represent the Cobb-Douglas function in Example 1 with $\beta = 1-\alpha$, i.e., $f(k) = Ak^{\alpha}$. The right panel of Fig. 2.3 shows a non-neoclassical case where only two alternative Leontief techniques are available, technique 1: $y = \min(A_1k, B_1)$, and technique 2: $y = \min(A_2k, B_2)$. In the exposed case it is assumed that $B_2 > B_1$ and $A_2 < A_1$ (if $A_2 \geq A_1$ at the same time as $B_2 > B_1$, technique 1 would not be efficient, because the same output could be obtained with less input of at least one of the factors by shifting to technique 2). If the available K and L are such that $k < B_1/A_1$ or $k > B_2/A_2$, some of either L or K, respectively, is idle. If, however, the available K and L are such that $B_1/A_1 < k < B_2/A_2$, it is efficient to *combine* the two techniques and use the fraction μ of K and L in technique 1 and the remainder in technique 2, where $\mu = (B_2/A_2 - k)/(B_2/A_2 - B_1/A_1)$. In this way we get the "labor productivity curve" OPQR (the envelope of the two

[7] See Math Tools.

[8] Given a Cobb-Douglas production function, both production factors are essential whether there is DRS, CRS, or IRS.

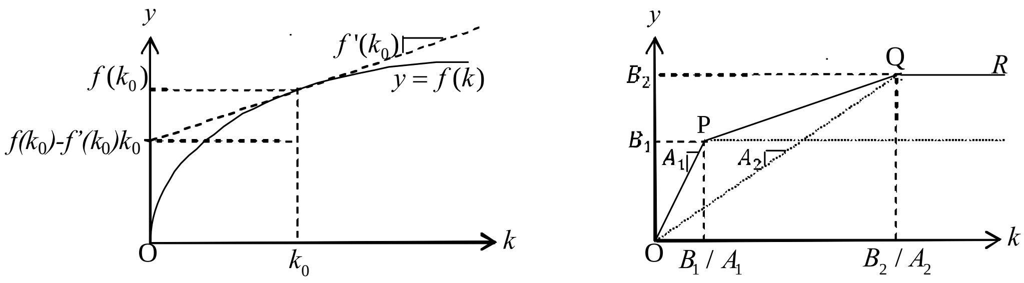

Figure 2.3: Two labor productivity curves based on CRS technologies. Left: neoclassical technology with Inada conditions for MPK satisfied. Right: a combination of two efficient Leontief techniques.

techniques) in Fig. 2.3. Note that for $k \rightarrow 0$, MPK stays equal to $A_1 < \infty$, whereas for all $k > B_2/A_2$, $MPK = 0$. A similar feature remains true, when we consider *many,* say n, alternative efficient Leontief techniques available. Assuming these techniques cover a considerable range wrt. the B/A ratios, we get a labor productivity curve looking more like that of a neoclassical CRS production function. On the one hand, this gives some intuition of what lies behind the assumption of a neoclassical CRS production function. On the other hand, it remains true that for all $k > B_n/A_n$, $MPK = 0$,[9] whereas for $k \rightarrow 0$, MPK stays equal to $A_1 < \infty$, thus questioning the lower Inada condition.

The implausibility of the lower Inada conditions is also underlined if we look at their implication in combination with the more reasonable upper Inada conditions. Indeed, the four Inada conditions taken *together* imply, under CRS, that output has no upper bound when either input goes to infinity for fixed amount of the other input (see Appendix C).

2.2 Technical change

When considering the movement over time of the economy, we shall often take into account the existence of *technical change*. When technical change occurs, the production function becomes time-dependent. Over time the production factors tend to become more productive: more output for given inputs. To put it differently: the isoquants move inward. When this is the case, we say that the technical change displays *technical progress*.

[9] Here we assume the techniques are numbered according to ranking with respect to the size of B.

A first step in taking this into account is to replace (2.1) by (2.4). In macroeconomics, however, technical change is often (and not too unrealistically) assumed to take the specific form of *Harrod-neutral technical change.*[10] This amounts to assuming we can write (2.4) in the form

$$Y_t = F(K_t, T_t L_t), \tag{2.19}$$

where F is a (time-independent) neoclassical production function, Y_t, K_t, and L_t are output, capital, and labor input, respectively, at time t, while T_t is the efficiency of labor and indicates the "technology level". Although one can imagine natural disasters implying a fall in T_t, generally T_t tends to rise over time and then we say that (2.19) represents *Harrod-neutral technical progress.* An alternative name for this is *labor-augmenting* technical progress (it acts *as if* the labor input were augmented).

If the function F in (2.19) is homogeneous of degree one (so that the technology for all t exhibits CRS wrt. capital and labor), we may write

$$\tilde{y}_t \equiv \frac{Y_t}{T_t L_t} = F(\frac{K_t}{T_t L_t}, 1) = F(\tilde{k}_t, 1) \equiv f(\tilde{k}_t), \qquad f' > 0, f'' < 0.$$

where $\tilde{k}_t \equiv K_t/(T_t L_t) \equiv k_t/T_t$ (habitually called the "effective" capital intensity or, if there is no risk of confusion, just the capital intensity). In accordance with a somewhat general trend in aggregate productivity data for industrialized countries, we often assume that T grows at a constant rate, g, so that, in discrete time $T_t = T_0(1+g)^t$ and in continuous time $T_t = T_0 e^{gt}$, where $g > 0$. The popularity in macroeconomics of the hypothesis of labor-augmenting technical progress derives from its consistency with Kaldor's "stylized facts", cf. Chapter 3.

There exists two alternative concepts of neutral technical progress. *Hicks-neutral* technical progress is said to occur if technological development is such that (2.4) can be written in the form

$$Y_t = T_t F(K_t, L_t), \tag{2.20}$$

where, again, F is a (time-independent) neoclassical production function, while T_t is the growing technology level.[11] The assumption of Hicks-neutrality has been used more in microeconomics and partial equilibrium analysis than in macroeconomics. Finally, *Solow-neutral* technical progress[12] is said to

[10] The name refers to the English economist Roy F. Harrod, 1900−1978.

[11] The name refers to the English economist and Nobel Prize winner John R. Hicks, 1904−1989.

[12] The name refers to the American economist and Nobel Prize winner Robert Solow (1924−).

occur if technological development is such that (2.4) can be written in the form

$$Y_t = F(T_t K_t, L_t). \tag{2.21}$$

Another name for the same is *capital-augmenting* technical progress (because the assumption is essentially that technical change acts as if the capital input were augmented).[13]

It is easily shown (Exercise 2.5) that the Cobb-Douglas production function satisfies all three neutrality criteria at the same time, if it satisfies one of them (which it does if technical change does not affect α and β, cf. (2.8) above). In Exercise 3.12 of the next chapter the reader is asked to show that within the class of neoclassical CRS production functions the Cobb-Douglas function is the only one with this property.

Note that the neutrality concepts do not say anything about the *source* of technical progress, only about the quantitative form in which it materializes. For instance, the occurrence of Harrod-neutrality, should not be interpreted as indicating that the technical change emanates specifically from the labor input in some sense.

There is another important taxonomy of technical change. We say that technical change is *embodied*, if taking advantage of new technical knowledge requires construction of new investment goods. The new technology is incorporated in the design of newly produced equipment, but this equipment will not participate in subsequent technical progress. An example: only the most recent vintage of a computer series incorporates the most recent advance in information technology. Then investment goods produced later (investment goods of a later "vintage") have higher productivity than investment goods produced earlier at the same resource cost. Thus investment becomes an important driving force in productivity increases.

We way formalize embodied technical progress by writing capital accumulation in the following way:

$$K_{t+1} - K_t = Q_t I_t - \delta K_t,$$

where I_t is gross investment in period t and Q_t measures the "quality" (productivity) of newly produced investment goods. The rising level of technology implies rising Q so that a given level of investment gives rise to a greater

[13] Macroeconomists' use of the value-laden term "technical progress" in connection with technical change may seem suspect. But the term should be interpreted as merely a label for certain types of shifts in isoquants in an abstract universe. At a more concrete and disaggregate level analysts of course make use of more refined notions about technical change, recognizing for example not only benefits of new technologies, but also the risks, including risk of fundamental mistakes (think of the introduction and later abandonment of ...).

and greater addition to the capital stock, K, measured in constant efficiency units. Even if technical change does not directly appear in the production function, that is, even if for instance (2.19) is replaced by $Y_t = F(K_t, L_t)$, the economy may thus experience a rising standard of living.

In contrast, *disembodied technical change* occurs when new technical knowledge increases the combined productivity of the production factors independently of when they were constructed or educated. If the K_t appearing in (2.19), (2.20), and (2.21) above refers the total, historically accumulated capital stock, then these expressions represent disembodied technical change. All vintages of the equipment benefit from a rise in the technology level T_t. No new investment is needed.

2.3 The concept of a representative firm

Many macroeconomic models make use of the simplifying notion of a *representative firm.* By this is meant a hypothetical firm whose production "represents" aggregate production (value added) in a sector or in society as a whole. Let n be the actual number of firms in the sector or in society and let Y_i, K_i, and L_i be output, capital input and labor input (per time unit) for firm i, $i = 1, 2, \ldots, n$. Further, let $Y = \Sigma_{i=1}^n Y_i$, $K = \Sigma_{i=1}^n K_i$, and $L = \Sigma_{i=1}^n L_i$. Ignoring technical change, suppose these aggregate variables in a given society turn out to be related through some production function, $F^*(\cdot)$, in the following way:

$$Y = F^*(K, L).$$

Then $F^*(K, L)$ is called the *aggregate production function* or the production function of the *representative* firm. It is *as if* aggregate production is the result of the behavior of such a single firm.

A simple example where the aggregate production function is well-defined is the following. Suppose that all firms have the same production function $F(\cdot)$ with CRS so that $Y_i = F(K_i, L_i)$, $i = 1, 2, \ldots, n$. In view of CRS,

$$Y_i = F(K_i, L_i) = L_i F(k_i, 1) \equiv L_i f(k_i),$$

where $k_i \equiv K_i/L_i$. Hence, facing given factor prices, all cost minimizing firms will choose the same capital intensity: $k_i = k$, for all i. From $K_i = kL_i$ then follows $\sum_i K_i = k \sum_i L_i$ so that $k = K/L$. Thence,

$$Y \equiv \sum Y_i = \sum L_i f(k_i) = f(k) \sum L_i = f(k)L = F(k, 1)L = F(K, L).$$

In this (trivial) case it is thus easy to construct an aggregate production function and this function turns out to be exactly the same as the (identical) CRS production functions of the individual firms.

Allowing for the existence of *different* output goods, capital goods, and technologies makes the issue more intricate, of course. Yet, if firms are price taking profit maximizers and there are nonincreasing returns to scale, then the aggregate outcome is *as if* the firms jointly maximized aggregate profit on the basis of their combined production technology. But the problem is that the conditions needed for an aggregate production function to be *well-behaved* (in the sense of inheriting simple properties from its constituent parts) are quite restrictive.[14] One aspect of the difficulties concerns the aggregation of the different kinds of equipment into one variable, the capital stock "K". In the 1960s there was a heated debate (the "Cambridge controversies") about these aggregation issues between a group of economists from Cambridge University, UK, and a group from Massachusetts Institute of Technology (MIT), which is in Cambridge, USA. The former group questioned the theoretical robustness of several of the neoclassical tenets, including the proposition that rising aggregate capital intensity tends to be associated with a falling rate of interest. Starting from the disaggregate level, an association of this sort is not a logical necessity because the relative prices tend to change, when the interest rate changes. While acknowledging this, the latter group maintained that in a macroeconomic context it is likely to cause devastating problems only under exceptional circumstances. In any event, since there is today no well-tried alternative, this book is about models that use aggregate constructs such as "K" (or "L" for that matter) as simplifying devices, assuming they are, for a broad class of cases, acceptable in a first approximation. It is another matter that when the role of imperfect competition is considered, we shall be ready to disaggregate the production side of the economy into several sectors producing different goods.

Like the representative firm the *representative household* is a simplifying notion that should be applied only when it does not get in the way of the issue to be studied. It is of course not a useful concept if we aim at understanding, say, the *interaction,* via financial intermediaries, between lending and borrowing households. Similarly, if the theme is conflicts of interests between firm owners and employees, the existence of *different* types of households have to be taken into account.

[14]Naturally, there are similar problems with the concept of an aggregate consumption function (in fact even more involved problems, in view of the role of individual budget constraints).